



Attorney Docket No. 21465-501 CIP2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS: Jonathan M. Rothberg, et al. CONF. No. 6233
SERIAL No.: 09/814,338 EXAMINER: Young J. Kim
FILING DATE: March 21, 2001 ART UNIT: 1637
FOR: METHOD OF SEQUENCING A NUCLEIC ACID

MAIL STOP AMENDMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF JONATHAN M. ROTHBERG, PH.D., OF PRIOR INVENTION

UNDER 37 C.F.R. §1.131

I, Jonathan M. Rothberg, Ph.D., of 454 Corporation and CuraGen Corporation declare and state that:

1. I am making this Declaration as the first named Inventor in the above-identified patent application ("the Application"). I am the Chairman of 454 Corporation and the founder, Chairman, and CEO of CuraGen Corporation. I earned a B.S. in Chemical Engineering from Carnegie Mellon University and a M.S., M. Phil., and Ph.D. in Biology from Yale University. I served as a Post-Doctoral Fellow at the Howard Hughes Medical Institute's Boyer Center for Molecular Medicine.
2. I am aware that in the Office Action dated July 28, 2004, the Examiner has cited U.S. Published Patent Application US 2002/0039728 ("Kain") under 35 U.S.C. §103(a) against claims 56-61, 64-68, 84-93, 96-100 of the Application. This declaration is being made to establish my invention of the claimed subject matter of a cavitated fiber optic wafer for a substrate and apparatus ("the Wafer") at a date prior to February 10, 2000, the effective filing date asserted by Kain.

3. I am the Inventor of the subject matter of claims 56-61, 64-68, 84-93, 96-100 of the Application. I hereby declare that I conceived of and constructively reduced to practice the claimed invention of the Wafer in the United States before February 10, 2000, the earliest priority date claimed by Kain.
4. I constructively reduced my invention to practice by filing U.S. Application No. 09/398,833 ("the '833 Application") on September 16, 1999, to which the current Application claims priority. A copy of the '833 Application is attached hereto as Exhibit 1. The Kain reference is attached as Exhibit 2. It is noted that the '833 Application has issued as U.S. Patent No. 6,274,320.
5. The '833 Application teaches a substrate for analyzing a nucleic acid (see, *inter alia*, p. 3, l. 23-25); a substrate comprising a cavitated fiber optic wafer (see, *inter alia*, p. 3, l. 26-30; p. 9, l. 22-25); individual optical fibers having a diameter between 3 μm and 100 μm (see, *inter alia*, p. 29, l. 4-5); and a bottom surface being highly polished to allow optical coupling to a second fused bundle of optical fibers (see, *inter alia*, p. 29, l. 25-27).
6. The '833 Application teaches a substrate comprising at least 10,000 reaction sites (see, *inter alia*, p. 12, l. 25-27; p. 37, l. 26-27; p. 45, l. 15-17); wells etched into the fiber optic wafer (p. 29, l. 17-20); a fiber optic wafer with a thickness of 1.5 mm (see, *inter alia*, p. 29, l. 1-3); a fiber optic wafer possessing handling properties similar to a plane of glass (see, *inter alia*, p. 29, l. 3-4); a fiber optic array the size of a cover slip (see, *inter alia*, p. 27, l. 20-22); and the depth of wells ranging from one half the diameter to three times the diameter of an individual optic fiber (see, *inter alia*, p. 29, l. 10-13).
7. The '833 Application teaches a cavitated fiber optic wafer with nucleic acids on the top surface (see, *inter alia*, p. 4, l. 5-7; original claims 59-62; Fig. 4); and nucleic acids immobilized on wells or beads (see, *inter alia*, p. 4, l. 5-7; p. 28, l. 26-27; p. 31, l. 1-2;

original claims 59-62; Fig. 4). The ‘833 Application teaches a plurality of microparticles with an anchored analytes, such as DNAs (see, *inter alia*, original claims 59-62); any material used as a solid support (see, *inter alia*, p. 9, l. 22-23); beads used as a solid support (see, *inter alia*, p. 31, l. 1-2); a solid support coupled to a bundle of optical fibers (see, *inter alia*, p. 28, l. 26-27).

8. The ‘833 Application teaches a plurality of beads having a pyrophosphate sequencing reagent attached thereto (see, *inter alia*, p. 9, l. 22-23; p. 26, l. 1-2; p. 28, l. 26-27; p. 31, l. 1-2). The ‘833 Application teaches pyrophosphate sequencing enzymes (e.g., sulfurylase, luciferase) immobilized to a solid support (see, *inter alia*, p. 26, l. 1-2); any material used as a solid support (see, *inter alia*, p. 9, l. 22-23); a solid support coupled to a bundle of optical fibers (see, *inter alia*, p. 28, l. 26-27); and beads used as a solid support (see, *inter alia*, p. 31, l. 1-2).
9. The ‘833 Application teaches nucleic acids separated by distances of approximately 10 μm to approximately 200 μm (see, *inter alia*, p. 4, l. 25-28); nucleic acids separated by distances of approximately 50 μm to approximately 150 μm (see, *inter alia*, p. 4, l. 25-28); nucleic acids separated by a distance of approximately 150 μm (see, *inter alia*, p. 4, l. 25-28); a wafer comprising 10^3 or more groups of nucleic acids (see, *inter alia*, original claims 41-43); a wafer comprising 10^4 or more groups of nucleic acids (see, *inter alia*, original claims 41-43); and a wafer comprising 10^5 or more groups of nucleic acids (see, *inter alia*, original claims 41-43).
10. The ‘833 Application teaches an apparatus for processing a plurality of nucleic acids (see, *inter alia*, p. 3, l. 23-25; p. 7, l. 8-11); a flow chamber with a cavitated fiber optic wafer disposed therein (see, *inter alia*, p. 28, l. 1-8); a fluid means for delivery of pyrophosphate sequencing reagents, including sequential delivery of dNTPs (see, *inter alia*, p. 24, l. 27-30); a detection means for detecting optical signals from the wells (see, *inter alia*, p. 26, l. 4-6); and a detection means comprising a CCD camera (see, *inter alia*, p. 26, l. 4-16).

11. As stated above, the '833 Application was filed on September 16, 1999. Accordingly, the teachings of the '833 Application demonstrate my conception and constructive reduction to practice of the claimed invention of the Wafer at least prior to September 16, 1999, which predates the earliest effective filing date asserted by Kain (i.e., February 10, 2000).
12. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by a fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.



Signature: Jonathan M. Rothberg

1/20/2005

Date

METHOD OF SEQUENCING A NUCLEIC ACID

FIELD OF THE INVENTION

The present invention relates to methods and apparatuses for determining the sequence of
5 a nucleic acid.

BACKGROUND OF THE INVENTION

Many diseases are associated with particular DNA sequences. The DNA sequences are often referred to as DNA sequence polymorphisms to indicate that the DNA sequence associated with a diseased state differs from the corresponding DNA sequence in non-afflicted individuals.
10 DNA sequence polymorphisms can include, e.g., insertions, deletions, or substitutions of nucleotides in one sequence relative to a second sequence. An example of a particular DNA sequence polymorphism is 5'-ATCG-3', relative to the sequence 5'-ATGG-3'. The first nucleotide 'G' in the latter sequence has been replaced by the nucleotide 'C' in the former sequence. The former sequence is associated with a particular disease state, whereas the latter sequence is found in individuals not suffering from the disease. Thus, the presence of the nucleotide sequence '5-ATCG-3' indicates the individual has the particular disease. This particular type of sequence polymorphism is known as a single-nucleotide polymorphism, or SNP, because the sequence difference is due to a change in one nucleotide.
15

Techniques which enable the rapid detection of as little as a single DNA base change is therefore important methodologies for use in genetic analysis. Because the size of the human genome is large, on the order of 3 billion base pairs, techniques for identifying polymorphisms must be sensitive enough to specifically identify the sequence containing the polymorphism in a potentially large population of nucleic acids.

Typically a DNA sequence polymorphism analysis is performed by isolating DNA from
25 an individual, manipulating the isolated DNA, e.g., by digesting the DNA with restriction enzymes and/or amplifying a subset of sequences in the isolated DNA. The manipulated DNA is then examined further to determine if a particular sequence is present.

Commonly used procedures for analyzing the DNA include electrophoresis. A common application of electrophoresis includes agarose or polyacrylamide gel electrophoresis. DNA
30 sequences are inserted, or loaded, on the gels and subjected to an electric field. Because DNA

carries a uniform negative charge, DNA will migrate through the gel based on a charge/mass ratio upon application of the electrical field. Smaller DNA molecules will migrate more rapidly through the gel than larger fragments. After electrophoresis has been continued for a sufficient length of time, the DNA molecules in the initial population of DNA sequences will have
5 separated according to their relative sizes.

Particular DNA molecules can then be detected using a variety of detection methodologies. For some applications, particular DNA sequences are identified by the presence of detectable tags, such as radioactive labels, attached to specific DNA molecules.

Electrophoretic-based separation analyses can be less desirable for applications in which
10 it is desirable to rapidly, economically, and accurately analyze a large number of nucleic acid samples for particular sequence polymorphisms. For example, electrophoretic-based analysis can require a large amount of input DNA. In addition, processing the large number of samples required for electrophoretic-based nucleic acid based analyses can be labor intensive.

Recently, automated electrophoresis systems have become available. However,
15 electrophoresis can be ill-suited for applications such as clinical sequencing, where relatively cost-effective units with high throughput are needed. Thus, the need for non-electrophoretic methods for sequencing is great. For many applications, electrophoreses is used in conjunction with DNA sequence analysis.

Several alternatives to electrophoretic-based sequencing have been described. These
20 include scanning tunnel electron microscopy, sequencing by hybridization, and single molecule detection methods.

Another alternative to electrophoretic-based separation is analysis is solid substrate-based nucleic acid analyses. These methods typically rely upon the use of large numbers of nucleic acid probes affixed to different locations on a solid support. These solid supports can include,
25 e.g., glass surfaces, plastic microtiter plates, plastic sheets, thin polymer, semi-conductors. The probes can be, e.g., adsorbed or covalently attached to the support, or can be microencapsulated or otherwise entrapped within a substrate membrane or film.

Substrate-based nucleic acid analyses can include applying a sample nucleic acid known or suspected of containing a particular sequence polymorphism to an array of probes attached to
30 the solid substrate. The nucleic acids in the population are allowed to hybridize to

complementary sequences attached to the substrate, if present. Hybridizing nucleic acid sequences are then detected in a detection step.

Solid support matrix-based hybridization and sequencing methodologies can require a high sample-DNA concentration and can be hampered by the relatively slow hybridization kinetics of nucleic acid samples with immobilized oligonucleotide probes. Often, only a small amount of template DNA is available, and it can be desirable to have high concentrations of the target nucleic acid sequence. Thus, substrate based detection analyses often include a step in which copies of the target nucleic acid, or a subset of sequences in the target nucleic acid, is amplified. Methods based on the Polymerase Chain Reaction (PCR), e.g., can increase a small number of probes targets by several orders of magnitude in solution. However, PCR can be difficult to incorporate into a solid-phase approach because the amplified DNA is not immobilized onto the surface of the solid support matrix.

Solid-phase based detection of sequence polymorphisms has been described. An example is a "mini-sequencing" protocol based upon a solid phase principle described by Hultman, et al., 1988. *Nucl. Acid. Res.* 17: 4937-4946; Syvanen, et al., 1990. *Genomics* 8: 684-692). In this study, the incorporation of a radiolabeled nucleotide was measured and used for analysis of a three-allelic polymorphism of the human apolipoprotein E gene. However, such radioactive methods are not well-suited for routine clinical applications, and hence the development of a simple, highly sensitive non-radioactive method for rapid DNA sequence analysis has also been of great interest.

SUMMARY OF THE INVENTION

The invention is based in part on the discovery of a highly sensitive method for determining the sequences of nucleic acids attached to solid substrates, and of novel substrate services for analyzing nucleic acid sequences.

Accordingly, in one aspect, the invention includes a substrate for analyzing a nucleic acid. The substrate includes a fiber optic surface onto which has been affixed one or more nucleic acid sequences. The fiber optic surface can be cavitated, e.g., a hemispherical etching of the opening of a fiber optic. The substrate can in addition include a plurality of bundled fiber optic surfaces, where one or more of the surfaces have anchored primers.

In another aspect, the invention includes an apparatus for analyzing a nucleic acid sequence. The apparatus can include a perfusion chamber, wherein the chamber includes a nucleic acid substrate, a conduit in communication with the perfusion chamber, an imaging system, *e.g.*, a fiber optic system, in communication with the perfusion chamber; and a data collection system in communication with the imaging system. The substrate can be a planar substrate. In other embodiments, the substrate can be the afore-mentioned fiber optic surface having nucleic acid sequences affixed to its termini.

In a further aspect, the invention includes a method for sequencing a nucleic acid. The method includes providing one or more or more nucleic acid anchor primers linked to a solid support and a plurality of circular nucleic acid templates. The nucleic acid anchor primer is then annealed to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex. The primed anchor primer-circular template complex is then combined with a polymerase to generate multiple copies of the circular nucleic acid template. Next, a sequencing primer is annealed to the circular nucleic acid template to yield a primed sequencing primer-circular nucleic acid template complex. The sequencing primer is then extended with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and a sequencing reaction byproduct, *e.g.*, inorganic pyrophosphate. If the predetermined nucleotide is incorporated into the primer, the sequencing reaction byproduct is generated and then identified, thereby determining the sequence of the nucleic acid. If desired, additional predetermined nucleotide triphosphates can be added, *e.g.*, sequentially, and the presence or absence of sequence byproducts associated with each reaction can be determined.

In a still further aspect, the invention includes a method for sequencing a nucleic acid by providing one or more nucleic acid anchor primers linked to a plurality of anchor primers linked to a fiber optic surface substrate, *e.g.*, the solid substrate discussed above.

In various embodiments of the apparatuses and methods described herein, the solid substrate includes two or more anchoring primers separated by approximately 10 μm to approximately 200 μm , 50 μm to approximately 150 μm , 100 μm to approximately 150 μm , or 150 μm . The solid support matrix can include a plurality of pads that are covalently linked to the solid support. The surface area of the pads can be, *e.g.*, 10 μm^2 and one or more pads can be

separated from one another by a distance ranging from approximately 50 μ m to approximately 150 μ m.

In preferred embodiments, at least a portion of the circular nucleic acid template is single-stranded DNA. The circular template can be, e.g., an open-circle nucleic acid or a closed circle 5 nucleic acid. The circular nucleic acid template can be, e.g., genomic DNA or RNA, or a cDNA copy thereof. The circular nucleic acid can be, e.g., 10-10,000 or 10-1000, 10-200, 10-100, 10-50, or 20-40 nucleotides in length.

In some embodiments, multiple copies of one or more circular nucleic acids in the population are generated by a polymerase chain reaction. In other embodiments, the primed 10 circular template is extended by rolling circle amplification (RCA) to yield a single-stranded concatamer of the annealed circular nucleic acid template. If desired, the template amplified by rolling circle amplification can be further amplified by annealing a reverse primer to the single-stranded concatamer to yield a primed concatamer template and combining the primed concatamer template with a polymerase enzyme to generate multiple copies of the concatamer 15 template. In still further embodiments, the template can be extended by a combination of PCR and RCA-amplification.

In preferred embodiments, sequencing byproduct analyzed is pyrophosphate. When pyrophosphate is used as the detected byproduct, a preferred nucleotide triphosphate for use by the polymerase in extending the primed sequencing primer is a dATP analog, e.g., adenosine 5'-phosphosulfate (APS).

Preferably, the pyrophosphate is detected by contacting the sequencing byproduct with ATP sulfurylase under conditions sufficient to form ATP. The ATP can then be detected, e.g., 20 with an enzyme which generates a detectable product upon reaction with ATP. A preferred enzyme for detecting the ATP is luciferase. If desired, a wash buffer, can be used between addition of various reactants herein. Preferably, apyrase is used to remove, e.g., unreacted dNTP 25 used to extend the sequencing primer. The wash buffer can optionally include apyrase.

The reactants and enzymes used herein, e.g., the ATP sulfurylase, luciferase, and apyrase, can be attached to the solid surface.

The anchor primer sequence can include, e.g. a biotin group, which can link the anchor 30 primer to the solid support via an avidin group attached to the solid support. In some

embodiments, the anchor primer is conjugated to a biotin-bovine serum albumin (BSA) moiety. The biotin-BSA moiety can be linked to an avidin-biotin group on the solid support. If desired, the biotin-BSA moiety on the anchor primer can be linked to a BSA group on the solid support in the presence of silane.

5 In some embodiments, the solid support includes at least one optical fiber.

The disclosures of one or more embodiments of the invention are set forth in the accompanying description below. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. Other features, objects, and advantages of 10 the invention will be apparent from the description and from the claims. In the specification and the appended claims, the singular forms include plural referents unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Unless expressly stated otherwise, the techniques employed or contemplated 15 herein are standard methodologies well known to one of ordinary skill in the art. The examples of embodiments are for illustration purposes only. All patents and publications cited in this specification are incorporated by reference.

20 **BRIEF DESCRIPTION OF THE DRAWINGS**

FIG. 1 is a schematic illustration of rolling circle based amplification using an anchor primer.

FIG. 2 is a drawing of a sequencing apparatus according to the present invention.

FIG. 3 is a drawing of a perfusion chamber according to the present invention.

25 FIG. 4 is a drawing of a cavitated fiber optic terminus of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The methods described herein include a sample preparation process in which multiple copies of individual single-stranded nucleic acid molecules, termed anchor primers, are linked to 30 a solid substrate. As is explained in more detail below, a region of the substrate containing at

least one linked anchor primer is a anchor pad. A plurality of anchor primers linked on a single solid surface can form an array.

A plurality of nucleic acid template sequences is then annealed to the array to form one or more primed circular templates. The primed circular templates are next amplified. After 5 amplification, a sequencing primer is annealed to the amplified nucleic acid and used to generate a sequencing product. The nucleotide sequence of the sequence product is then determined, thereby allowing for the determination of the nucleic acid.

The methods and apparatuses described herein allow for the determination of nucleic acid sequence information without the need for first cloning a nucleic acid. In addition, the method is 10 highly sensitive and can be used to determine the nucleotide sequence of a template nucleic acid which is present in only a few copies in a starting population of nucleic acids.

The methods and apparatuses described are generally useful for any application which the identification of any particular nucleic acid sequence is desired. For example, the methods allow 15 for identification of single nucleotide polymorphisms (SNPs) and transcript profiling. Other uses include sequencing of artificial DNA constructs to confirm or elicit their primary sequence, or to isolate specific mutant clones from random mutagenesis screens, as well as to obtain the sequence of cDNA from single cells, whole tissues or organisms from any developmental stage or environmental circumstance in order to determine the gene expression profile from that specimen. In addition, the methods allow for the sequencing of PCR products and/or cloned 20 DNA fragments of any size isolated from any source.

The methods of the present invention can be also used for the sequencing of DNA fragments generated by analytical techniques that probe higher order DNA structure by their differential sensitivity to enzymes, radiation or chemical treatment (*e.g.*, partial DNase treatment of chromatin), or for the determination of the methylation status of DNA by comparing sequence 25 generated from a given tissue with or without prior treatment with chemicals that convert methyl-cytosine to thymine (or other nucleotide) as the effective base recognized by the polymerase. Further, the methods of the present invention can be used to assay cellular physiology changes occurring during development or senescence at the level of primary sequence.

Methods of Sequencing Nucleic Acids

Structure of Anchor Primers

Anchor primers in general include a stalk region and at least two contiguous adapter regions. The stalk region is present at the 5' end of the anchor primer and includes a region of nucleotides for attaching the anchor primer to the solid substrate.

The anchor primer in general includes a region which hybridizes to a complementary sequence present in one or more members of a population of nucleic acid sequences. In some embodiments, the anchor primer includes two adjoining regions which hybridize to complementary regions ligated to separate ends of a target nucleic acid sequence. This embodiment is illustrated in FIG.1, which is discussed in more detail below.

In some embodiments, the adapter regions in the anchor primers are complementary to non-contiguous regions of sequence present in a second nucleic acid sequence. Each adapter region, for example, can be homologous to each terminus of a fragment produced by digestion with one or more restriction endonucleases. The fragment can include, e.g., a sequence known or suspected to contain a sequence polymorphism.

In another example, the anchor primer may contain two adapter regions that are homologous to a gapped, i.e., non-contiguous because of a deletion of one or more nucleotides, region of a target nucleic acid sequence. For example, e.g., a target sequence in population of nucleic acids sequences. When adapter regions having these sequences are used, an aligning oligonucleotide corresponding to the gapped sequence may be annealed to the anchor primer along with a population of template nucleic acid molecules.

The anchor primer may optionally contain additional elements, e.g., one or more restriction enzyme recognition sites, RNA polymerase binding sites (e.g., a T7 promoter site).

One or more of the adapter regions may include, e.g., a restriction enzyme recognition site or sequences present in identified DNA sequences, e.g., sequences present in known genes. One or more adapter regions may also include sequences known to flank sequence polymorphisms. Sequence polymorphisms include nucleotide substitutions, insertions, deletions, or other rearrangements which result in a sequence difference between two otherwise identical

nucleic acid sequences. An example of a sequence polymorphism is a single nucleotide polymorphism (SNP).

Linking of Anchor Primers to a Solid Support

5 In general, any nucleic acid capable of base-pairing can be used as an anchor primer. In some embodiments, the anchor primer is an oligonucleotide. As utilized herein the term oligonucleotide includes linear oligomers of natural or modified monomers or linkages, e.g., deoxyribonucleosides, ribonucleosides, anomeric forms thereof, peptide nucleic acids (PNAs), and the like, that are capable of specifically binding to a target polynucleotide by way of a
10 regular pattern of monomer-to-monomer interactions. These types of interactions can include, e.g., Watson-Crick type of base-pairing, base stacking, Hoogsteen or reverse-Hoogsteen types of base-pairing, or the like. Generally, the monomers are linked by phosphodiester bonds, or analogs thereof, to form oligonucleotides ranging in size from, e.g., 3-200, 8-150, 10-100, 20-
15 80, or 25-50 monomeric units. Whenever an oligonucleotide is represented by a sequence of letters, it is understood that the nucleotides are oriented in the 5' → 3' direction, from left-to-right, and that the letter "A" denotes deoxyadenosine, the letter "T" denotes thymidine, the letter "C" denotes deoxycytosine, and the letter "G" denotes deoxyguanosine, unless otherwise noted herein. The oligonucleotides of the present invention can include non-natural nucleotide analogs.
20 However, where, for example, processing by enzymes is required, or the like, oligonucleotides comprising naturally-occurring nucleotides are generally required for maintenance of biological function.

25 Any material can be used as the solid support material, as long as the surface allows for stable attachment of the primers and detection of nucleic acid sequences. The solid support material can be planar or can be cavitated, e.g., in a cavitated terminus of a fiber optic. In some embodiments, the solid support is optically transparent, e.g., glass.

30 The anchor primer can be linked to the solid support to reside on or within the solid support. In some embodiments, the plurality of anchor primers is linked to the solid support so they are spaced regular intervals within an array. The periodicity between primers is preferably greater than either the diffusion rate of the products of the sequencing reactions or the optical resolving power of the detection system, both of which are described in more detail below. The

distance between primers on a solid substrate can be, 10-400 μm , 50-150 μm , 100-150 μm , or 150 μm .

An array of attachment sites on the optically transparent solid support is constructed using lithographic techniques commonly used in the construction of electronic integrated circuits 5 as described in, e.g., techniques for attachment described in U.S. Patent Nos. 5,5143,854, 5,445,934, 5,744,305, and 5, 800,992; Chee *et al.*, *Science* 274: 610-614 (1996); Fodor *et al.*, *Nature* 364: 555-556 (1993); Fodor *et al.*, *Science* 251: 767-773 (1991); Gushin, *et al.*, *Anal. Biochem.* 250: 203-211 (1997); Kinosita *et al.*, *Cell* 93: 21-24 (1998); Kato-Yamada *et al.*, *J. Biol. Chem.* 273: 19375-19377 (1998); and Yasuda *et al.*, *Cell* 93: 1117-1124 (1998).

10 Photolithography and electron beam lithography sensitize the solid support or substrate with a linking group that allows attachment of a modified biomolecule (e.g., proteins or nucleic acids). See e.g., Service, *Science* 283: 27-28 (1999); Rai-Choudhury, HANDBOOK OF MICROLITHOGRAPHY, MICROMACHINING, AND MICROFABRICATION, VOLUME I: MICROLITHOGRAPHY, Volume PM39, SPIE Press (1997). Alternatively, an array of sensitized 15 sites can be generated using thin-film technology as described in Zasadzinski *et al.*, *Science* 263: 1726-1733 (1994). The contents of all of these patents and publications are incorporated by reference in their entirety.

Anchor primers are linked to the solid substrate at the sensitized sites. A region of a solid substrate containing a linked primer is an anchor pad. Thus, by specifying the sensitized 20 states on the solid support, it is possible to form an array or matrix of anchored pads. The anchor pads can, e.g., small diameter spots etched at evenly spaced intervals on the solid support.

The anchor primer can be attached to the solid support via a covalent or non-covalent interaction. Examples of such linkages common in the art include Ni²⁺/hexahistidine, streptavidin/biotin, avidin/biotin, glutathione S-transferase (GST)/glutathione, monoclonal 25 antibody/antigen, and maltose binding protein/maltose. Samples containing the appropriate tag are incubated with the sensitized substrate so that a single molecule attaches at each sensitized site.

The biotin-(strept-)avidin methodology provides several different ways to immobilize the anchor on the solid support. One biotin-(strept-)avidin-based anchoring method uses a thin layer 30 of a photoactivatable biotin analog dried onto a solid surface. (Hengsakul and Cass, 1996.

Bioconjugate Chem. 7: 249-254). The biotin analog is then exposed to white light through a mask, so as to create defined areas of activated biotin. Avidin (or streptavidin) is then added and allowed to bind to the activated biotin. The avidin possesses free biotin binding sites which can be utilized to "anchor" the biotinylated oligonucleotides through a biotin-(strept-)avidin linkage.

5 Alternatively, the anchor primer can be attached to the solid support with a biotin derivative possessing a photo-removable protecting group. This moiety is covalently bound to bovine serum albumin (BSA), which is attached to the solid support, e.g., a glass surface. See Pirrung and Huang, 1996. *Bioconjugate Chem.* 7: 317-321. A mask is then used to create activated biotin within the defined irradiated areas. Avidin may then be localized to the
10 irradiated area, with biotinylated DNA subsequently attached through a BSA-biotin-avidin-biotin link. If desired, an intermediate layer of silane is deposited in a self-assembled monolayer on a silicon dioxide silane surface that can be patterned to localize BSA binding in defined regions. See e.g., Mooney, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 12287-12291.

15 Each sensitized site on a solid support is potentially capable of attaching multiple anchor primers. Thus, each anchor pad may include one or more anchor primers. It is preferable to maximize the number of pads that have only a single productive reaction center (e.g., the number of pads that, after the extension reaction, have only a single sequence extended from the anchor primer). This can be accomplished by techniques which include, but are not limited to: (i) varying the dilution of biotinylated anchor primers that are washed over the surface; (ii) varying
20 the incubation time that the biotinylated primers are in contact with the avidin surface; or (iii) varying the concentration of open- or closed-circular template so that, on average, only one primer on each pad is extended to generate the sequencing template.

25 In some embodiments, each individual pad contains just one linked anchor primer. Pads having only one anchor primer can be made by performing limiting dilutions of a selected anchor primer on to the solid support such that, on average, only one anchor primer is deposited on each pad. The concentration of anchor primer to be applied to a pad can be calculated utilizing, for example, a Poisson distribution model.

30 In order to maximize the number of reaction pads that contain a single anchor primer, a series of dilution experiments are performed in which a range of anchor primer concentrations or circular template concentrations are varied. For highly dilute concentrations of primers, primers

and circular templates binding to the same pad will be independent of each other, and a Poisson distribution will characterize the number of anchor primers extended on any one pad. Although there will be variability in the number of primers that are actually extended, a maximum of 37% of the pads will have a single extended anchor primer (the number of pads with a single anchor oligonucleotide). This number can be obtained as follows.

Let N_p be the average number of anchor primers on a pad and f be the probability that an anchor primer is extended with a circular template. Then the average number of extended anchor primers per pad is $N_p f$, which is defined as the quantity a . There will be variability in the number of primers that are actually extended. In the low-concentration limit, primers and circular templates binding to the same pad will be independent of each other, and a Poisson distribution $P(n)$ will characterize the number of anchor primers n extended on any pad. This distribution may be mathematically defined by: $P(n) = (a^n / n!) \exp(-a)$, with $P(1) = a \exp(-a)$. The probability $P(1)$ assumes its maximum value $\exp(-1)$ for $a = 1$, with 37% of pads having a single extended anchor primer.

A range of anchor primer concentrations and circular template concentrations may be subsequently scanned to find a value of $N_p f$ closest to 1. A preferable method to optimize this distribution is to allow multiple anchor primers on each reaction pad, but use a limiting dilution of circular template so that, on average, only one primer on each pad is extended to generate the sequencing template.

Alternatively, at high concentration of anchor primers, multiple anchor primers will likely be bound on each reaction pad, but a limiting dilution of circular template may be used so that, on average, only one primer on each pad anneals to a template molecule and is extended to amplify the sequencing template.

Where the reaction pads are arrayed on a planar surface, the individual pads are approximately 10 μm on a side, with a 100 μm spacing between adjacent pads. Hence, on a 1 cm surface a total of approximately 10,000 pads could be deposited, and, according to the Poisson distribution, approximately 3700 of these will contain a single anchor primer. In certain embodiments, after the primer oligonucleotide has been attached to the solid support, modified, e.g., biotinylated, enzymes are deposited to bind to the remaining, unused avidin binding sites on the planar surface.

In other embodiments multiple anchor primers are attached to any one individual pad in an array. Limiting dilutions of a plurality of circular nucleic acid templates (described in more detail below) may be hybridized to the anchor primers so immobilized such that, on average, only one primer on each pad is hybridized to a nucleic acid template. Library concentrations to be used may be calculated utilizing, for example, limiting dilutions and a Poisson distribution model.

Libraries of single-stranded circular templates

A plurality of nucleic acid templates, *e.g.*, a nucleic acid library, in general includes open circular or closed circular nucleic acid molecules. A “closed circle” is a covalently closed circular nucleic acid molecule, *e.g.*, a circular DNA or RNA molecule. An “open circle” is a linear single-stranded nucleic acid molecule having a 5’ phosphate group and a 3’ hydroxyl group. The ends of a given open circle nucleic acid molecule can be ligated by DNA ligase. Sequences at the 5’ and 3’ ends of the open circle molecule are complementary to two regions of adjacent nucleotides in a second nucleic acid molecule, *e.g.*, an adapter region of an anchor primer, or to two regions that are nearly adjoining in a second DNA molecule. Thus, the ends of the open-circle molecule can be ligated using DNA ligase, or extended by DNA polymerase in a gap-filling reaction. Open circles are described in detail in Lizardi, U.S. Pat. No. 5, 854,033. An open circle can be converted to a closed circle in the presence of a DNA ligase (for DNA) or RNA ligase following, *e.g.*, annealing of the open circle to an anchor primer.

Circularized oligonucleotide probes (*i.e.*, padlock probes) are comprised of two target sequence-complementarity sequences which are connected by a linker which may possess detectable functionalities. The linkers can be ligated to ends of members of a library of nucleic acid sequences that have been, *e.g.*, physically sheared or digested with restriction endonucleases.

The 5'- and 3'-terminal regions of these linear oligonucleotides are designed to basepair adjacent to one another on a specific target sequence strand, thus the termini of the linear oligonucleotide are brought into juxtaposition by hybridization to the target sequence. This juxtaposition allows the two probe segments (if properly hybridized) to be covalently-bound by enzymatic ligation (*e.g.*, with T₄ DNA ligase), thus converting the probes to circularly-closed

molecules which are catenated to the specific target sequences (see e.g., Nilsson, *et al.*, 1994. *Science* **265**: 2085-2088). The resulting probes are suitable for the simultaneous analysis of many gene sequences both due to their specificity and selectivity for gene sequence variants (see e.g., Lizardi, *et al.*, 1998. *Nat. Genet.* **19**: 225-232; Nilsson, *et al.*, 1997. *Nat. Genet.* **16**: 252-255) and due to the fact that the resulting reaction products remain localized to the specific target sequences. Moreover, intramolecular ligation of many different probes is expected to be less susceptible to non-specific cross-reactivity than multiplex PCR-based methodologies where non-cognate pairs of primers can give rise to irrelevant amplification products (see e.g., Landegren and Nilsson, 1997. *Ann. Med.* **29**: 585-590).

10 The starting library can be either single-stranded or double-stranded, as long as it includes a region that, if present in the library, is available for annealing, or can be made available for annealing, to an anchor primer sequence.

Library templates can include multiple elements, including, but not limited to, one or more regions that are complementary to the anchor primer. For example, the template libraries 15 may include a region complementary to a sequencing primer, a control nucleotide region, and an insert sequence comprised of the sequencing template to be subsequently characterized. As is explained in more detail below, the control nucleotide region is used to calibrate the relationship between the amount of byproduct and the number of nucleotides incorporated. As utilized herein the term "complement" refers to nucleotide sequences that are able to hybridize to a specific 20 nucleotide sequence to form a matched duplex.

In one embodiment, a library template includes: (i) two distinct regions that are complementary to the anchor primer, (ii) one region complementary to the sequencing primer, (iii) one control nucleotide region, (iv) an insert sequence of 30 - 100 nucleotides that is to be sequenced. The template can, of course, include two, three, or all four of these features.

25 The template nucleic acid can be constructed from any source of nucleic acid, e.g., any cell, tissue, or organism, and can be generated by any art-recognized method. Suitable methods include, e.g., sonication of genomic DNA and digestion with one or more restriction endonucleases (RE) to fragment a population of nuclei acid molecules, e.g., genomic DNA . Preferably, one or more of the restriction enzymes have distinct four-base recognition sequences..

30 Examples of such enzymes include, e.g., Sau3A1, MspI, and TaqI. Preferably, the enzymes are

used in conjunction with anchor primers having regions containing recognition sequences for the corresponding restriction enzymes. In some embodiments, the one or both adapter regions anchor primers contain additional sequences adjoining known restriction enzyme recognition sequences, thereby allowing for capture or annealing of specific restriction fragments of interest
5 to the anchor primer.

In other embodiments, the restriction enzyme is used with a type IIS restriction enzyme.

Alternatively, template libraries can be made by generating a complementary DNA (cDNA) library from RNA, *e.g.*, messenger RNA (mRNA). The cDNA library can, if desired, be further processed with restriction endonucleases to obtain either 3' signature sequences, internal
10 fragments, or 5' fragments. adapter regions in the anchor primer, libraries containing a sequence of interest, *e.g.*, a known or suspected sequence polymorphism on a restriction fragment.

Annealing and Amplification of Primer-Template Nucleic Acid Complexes

Libraries of nucleic acids are annealed to anchor primer sequences using recognized
15 techniques (see, *e.g.*, Hatch, *et al.*, 1999. *Genet. Anal. Biomol. Engineer.* 15: 35-40; Kool, U.S. Patent No. 5,714,320 and Lizardi, U.S. Patent No. 5,854,033). In general, any procedure for annealing the anchor primers to the template nucleic acid sequences is suitable as long as it results in formation of specific, *i.e.*, perfect or nearly perfect, complementarity between the adapter region or regions in the anchor primer sequence and a sequence present in the template
20 library.

A number of *in vitro* nucleic acid amplification techniques may be utilized to extend the anchor primer sequence. The size of the amplified DNA should be smaller than the size of the anchor pad and also smaller than the distance between anchoring pads.

The amplification is typically performed in the presence of a polymerase, *e.g.*, a DNA or
25 RNA-directed DNA polymerase, and one, two, three, or four types of nucleotide triphosphates, and, optionally, auxiliary binding proteins. In general, any polymerase capable of extending a primed 3'-OH group can be used as long as it lacks a 3' to 5' exonuclease activity. Suitable polymerases include, *e.g.*, the DNA polymerases from *Bacillus stearothermophilus*, *Thermus aquaticus*, *Pyrococcus furiosis*, *Thermococcus litoralis*, and *Thermus thermophilus*,
30 bacteriophage T₄ and T₇, and the *E. coli* DNA polymerase I Klenow fragment. Suitable RNA-

directed DNA polymerases include, e.g., the reverse transcriptase from the Avian Myeloblastosis Virus, the reverse transcriptase from the Moloney Murine Leukemia Virus, and the reverse transcriptase from the Human Immunodeficiency Virus-I.

A number of *in vitro* nucleic acid amplification techniques have been described. These amplification methodologies may be differentiated into those methods: (i) which require temperature cycling - polymerase chain reaction (PCR) (see e.g., Saiki, *et al.*, 1995. *Science* **230**: 1350-1354), ligase chain reaction (see e.g., Barany, 1991. *Proc. Natl. Acad. Sci. USA* **88**: 189-193; Barringer, *et al.*, 1990. *Gene* **89**: 117-122) and transcription-based amplification (see e.g., Kwok, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* **86**: 1173-1177) and (ii) isothermal amplification systems - self-sustaining, sequence replication (see e.g., Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* **87**: 1874-1878); the Q β replicase system (see e.g., Lizardi, *et al.*, 1988. *BioTechnology* **6**: 1197-1202); strand displacement amplification Nucleic Acids Res. 1992 Apr 11;20(7):1691-6.; and the methods described in PNAS 1992 Jan 1;89(1):392-6; and NASBA J Virol Methods. 1991 Dec;35(3):273-86.

Isothermal amplification also includes rolling circle-based amplification (RCA). RCA is discussed in, e.g., Kool, U.S. Patent No. 5,714,320 and Lizardi, U.S. Patent No. 5,854,033; Hatch, *et al.*, 1999. *Genet. Anal. Biomol. Engineer.* **15**: 35-40. The result of the RCA is a single DNA strand extended from the 3' terminus of the anchor primer (and thus is linked to the solid support matrix) and including a concatamer containing multiple copies of the circular template annealed to a primer sequence. Typically, 10,000 or more copies of circular templates, each having a size of approximately 100 nucleotides size range, can be obtained with RCA.

The product of RCA amplification following annealing of a circular nucleic acid molecule to an anchor primer is shown schematically in FIG. 1. A circular template nucleic acid 102 is annealed to an anchor primer 104, which has been linked to a surface 106 at its 5' end and has a free 3' OH available for extension. The circular template nucleic acid 102 includes two adapter regions 108 and 110 which are homologous to regions of sequence in the anchor primer 104. Also included in the circular template nucleic acid 102 is an insert 112 and a region 114 homologous to a sequencing primer, which is used in the sequencing reactions described below.

Upon annealing, the free 3'-OH on the anchor primer 104 can be extended using sequences within the template nucleic acid 102. The anchor primer 102 can be extended along

the template multiple times , with each iteration adding to the sequence extended from the anchor primer a sequence complementary to the circular template nucleic acid. Four iterations, or four rounds of rolling circle replication, are shown in FIG.1 as the extended anchor primer amplification product 114. Extension of the anchor primer results in an amplification product 5 covalently attached to the substrate 106.

Circular oligonucleotides which are generated during polymerase-mediated DNA replication are dependent upon the relationship between the template and the site of replication initiation. In double-stranded DNA templates, the critical features include whether the template is linear or circular in nature, and whether the site of initiation of replication (*i.e.*, the replication 10 “fork”) is engaged in synthesizing both strands of DNA or only one. In conventional double-stranded DNA replication, the replication fork is treated as the site at which the new strands of DNA are synthesized. However, in linear molecules (whether replicated unidirectionally or bidirectionally), the movement of the replication fork(s) generate a specific type of structural motif. If the template is circular, one possible spatial orientation of the replicating molecule. 15 takes the form of an θ structure.

Alternatively, RCA can occur when the replication of the duplex molecule begins at the origin. Subsequently, a nick opens one of the strands, and the free 3'-terminal hydroxyl moiety generated by the nick is extended by the action of DNA polymerase. The newly synthesized strand eventually displaces the original parental DNA strand. This aforementioned type of 20 replication is known as rolling-circle replication (RCR) because the point of replication may be envisaged as “rolling around” the circular template strand and, theoretically, it could continue to do so indefinitely. As it progresses, the replication fork extends the outer DNA strand the previous partner. Additionally, because the newly synthesized DNA strand is covalently-bound 25 to the original template, the displaced strand possesses the original genomic sequence (*e.g.*, gene or other sequence of interest) at its 5'-terminus. In rolling-circle replication, the original genomic sequence is followed by any number of “replication units” complementary to the original template sequence, wherein each replication unit is synthesized by continuing revolutions of said original template sequence. Hence, each subsequent revolution displaces the DNA which is synthesized in the previous replication cycle.

In vivo, rolling-circle replication is utilized in several biological systems. For example, in certain bacteriophage, their genome consists of single-stranded, circular DNA. During replication, the circular DNA is initially converted to a duplex form, which is then replicated by the aforementioned rolling-circle replication mechanism. The displaced terminus generates a series of genomic units, which can be cleaved and inserted into the phage particles, or they can be utilized for further replication cycles by the phage. Additionally, the displaced single-strand of a rolling-circle can be converted to duplex DNA by synthesis of a complementary DNA strand. This synthesis can be used to generate the concatemeric duplex molecules required for the maturation of certain phage DNAs. For example, this provides the principle pathway by which λ bacteriophage matures. Rolling-circle replication is also used *in vivo* to generate amplified rDNA in *Xenopus* oocytes, and this fact may help explain why the amplified rDNA is comprised of a large number of identical repeating units. In this case, a single genomic repeating unit is converted into a rolling-circle. The displaced terminus is then converted into duplex DNA which is subsequently cleaved from the circle so that the two termini can be ligated together so as to generate the amplified circle of rDNA.

Through the use of the RCR reaction, a strand may be generated which represents many tandem copies of the complement to the circularized molecule. For example, RCR has recently been utilized to obtain an isothermal cascade amplification reaction of circularized padlock probes *in vitro* in order to detect single-copy genes in human genomic DNA samples (see Lizardi, *et al.*, 1998. *Nat. Genet.* **19**: 225-232). In addition, RCR has also been utilized to detect single DNA molecules in a solid phase-based assay, although difficulties arose when this technique was applied to *in situ* hybridization (see Lizardi, *et al.*, 1998. *Nat. Genet.* **19**: 225-232).

The development of a method of amplifying short DNA molecules which have immobilized to a solid support, termed rolling circle amplification (RCA) has been recently described in the literature (see *e.g.*, Hatch, *et al.*, 1999. Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection. *Genet. Anal. Biomol. Engineer.* **15**: 35-40; Zhang, *et al.*, 1998. Amplification of target-specific, ligation-dependent circular probe. *Gene* **211**: 277-85; Baner, *et al.*, 1998. Signal amplification of padlock probes by rolling circle replication. *Nucl. Acids Res.* **26**: 5073-5078; Liu, *et al.*, 1995. Rolling

circle DNA synthesis: small circular oligonucleotides as efficient templates for DNA polymerase. *J. Am. Chem. Soc.* **118**: 1587-1594; Fire and Xu, 1995. Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. USA* **92**: 4641-4645; Nilsson, *et al.*, 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**: 2085-2088). RCA 5 targets specific DNA sequences through hybridization and a DNA ligase reaction. The circular product is then subsequently used as a template in a rolling circle replication reaction.

Rolling-circle amplification (RCA) driven by DNA polymerase can replicate circularized oligonucleotide probes with either linear or geometric kinetics under isothermal conditions. In the presence of two primers (one hybridizing to the + strand, and the other, to the - strand of 10 DNA), a complex pattern of DNA strand displacement ensues which possesses the ability to generate 1×10^9 or more copies of each circle in a short period of time (*i.e.*, less-than 90 minutes), enabling the detection of single-point mutations within the human genome. Using a single primer, RCA generates hundreds of randomly-linked copies of a covalently closed circle in several minutes. If solid support matrix-associated, the DNA product remains bound at the site 15 of synthesis, where it may be labeled, condensed, and imaged as a point light source. For example, linear oligonucleotide probes, which can generate RCA signals, have been bound covalently onto a glass surface. The color of the signal generated by these probes indicates the allele status of the target, depending upon the outcome of specific, target-directed ligation events. As RCA permits millions of individual probe molecules to be counted and sorted, it is 20 particularly amenable for the analysis of rare somatic mutations. RCA also shows promise for the detection of padlock probes bound to single-copy genes in cytological preparations.

In addition, a solid-phase RCA methodology has also been developed to provide an effective method of detecting constituents within a solution. Initially, a recognition step is used to generate a complex consisting of a DNA primer duplexed with a circular template is bound to 25 a surface. A polymerase enzyme is then used to amplify the bound complex. RCA uses small DNA probes that are amplified to provide an intense signal using detection methods, including the methods described in more detail below.

Other examples of isothermal amplification systems include, *e.g.*, (*i*) self-sustaining, sequence replication (see *e.g.*, Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* **87**: 1874-1878),

(ii) the Q β replicase system (see e.g., Lizardi, et al., 1988. *BioTechnology* 6: 1197-1202), and
(iii) nucleic acid sequence-based amplification (NASBATM; see Kievits, et al., 1991. *J. Virol. Methods* 35: 273-286).

Determining the nucleotide sequence of the sequence product

5 Amplification of a nucleic acid template as described above results in multiple copies of a template nucleic acid sequence covalently linked to an anchor primer. In one embodiment, a region of the sequence product is determined by annealing a sequencing primer to region of the template nucleic acid, and then contacting the sequencing primer with a DNA polymerase and a known nucleotide triphosphate, i.e., dATP, dCTP, dGTP, dTTP, or an analog of one of these
10 nucleotides.

The sequence primer can be any length or base composition, as long as it is capable of specifically annealing to a region of the amplified nucleic acid template. No particular structure is required for the sequencing primer is required so long as it is able to specifically prime a region on the amplified template nucleic acid. Preferably, the sequencing primer is
15 complementary to a region of the template that is between the sequence to be characterized and the sequence hybridizable to the anchor primer. The sequencing primer is extended with the DNA polymerase to form a sequence product. The extension is performed in the presence of one or more types of nucleotide triphosphates, and if desired, auxiliary binding proteins.

Incorporation of the dNTP is determined by assaying for the presence of a sequencing
20 byproduct. In a preferred embodiment, the nucleotide sequence of the sequencing product is determined by measuring inorganic pyrophosphate (PPi) liberated from a nucleotide triphosphate (dNTP) as the NTP is incorporated into an extended sequence primer. This method of sequencing, termed PyrosequencingTM technology (PyroSequencing AB, Stockholm, Sweden) can be performed in solution (liquid phase) or as a solid phase technique. PPi-based sequencing
25 methods are described generally in, e.g., WO9813523A1, Ronaghi, et al., 1996. *Anal. Biochem.* 242: 84-89, and Ronaghi, et al., 1998. *Science* 281: 363-365 (1998). These disclosures of PPi sequencing are incorporated herein in their entirety, by reference.

Pyrophosphate released under these conditions can be detected enzymatically (e.g., by the generation of light in the luciferase-luciferin reaction). Such methods enable a nucleotide to be

identified in a given target position, and the DNA to be sequenced simply and rapidly while avoiding the need for electrophoresis and the use of potentially dangerous radiolabels.

PPi can be detected by a number of different methodologies, and various enzymatic methods have been previously described (see e.g., Reeves, *et al.*, 1969. *Anal. Biochem.* **28**: 282-5 287; Guillory, *et al.*, 1971. *Anal. Biochem.* **39**: 170-180; Johnson, *et al.*, 1968. *Anal. Biochem.* **15**: 273; Cook, *et al.*, 1978. *Anal. Biochem.* **91**: 557-565; and Drake, *et al.*, 1979. *Anal. Biochem.* **94**: 117-120).

PPi liberated as a result of incorporation of a dNTP by a polymerase can be converted to ATP using, e.g., an ATP sulfurylase. This enzyme has been identified as being involved in sulfur metabolism. Sulfur, in both reduced and oxidized forms, is an essential mineral nutrient for plant and animal growth (see e.g., Schmidt and Jager, 1992. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **43**: 325-349). In both plants and microorganisms, active uptake of sulfate is followed by reduction to sulfide. As sulfate has a very low oxidation/reduction potential relative to available cellular reductants, the primary step in assimilation requires its activation via an ATP-dependent reaction (see e.g., Leyh, 1993. *Crit. Rev. Biochem. Mol. Biol.* **28**: 515-542). ATP sulfurylase (ATP: sulfate adenylyltransferase; EG 2.7.7.4) catalyzes the initial reaction in the metabolism of inorganic sulfate (SO_4^{2-}); see e.g., Robbins and Lipmann, 1958. *J. Biol. Chem.* **233**: 686-690; Hawes and Nicholas, 1973. *Biochem. J.* **133**: 541-550) In this reaction SO_4^{2-} is activated to adenosine 5'-phosphosulfate (APS).

ATP sulfurylase has been highly purified from several sources, such as *Saccharomyces cerevisiae* (see e.g., Hawes and Nicholas, 1973. *Biochem. J.* **133**: 541-550); *Penicillium chrysogenum* (see e.g., Renosto, *et al.*, 1990. *J. Biol. Chem.* **265**: 10300-10308); rat liver (see e.g., Yu, *et al.*, 1989. *Arch. Biochem. Biophys.* **269**: 165-174); and plants (see e.g., Shaw and Anderson, 1972. *Biochem. J.* **127**: 237-247; Osslund, *et al.*, 1982. *Plant Physiol.* **70**: 39-45). Furthermore, ATP sulfurylase genes have been cloned from prokaryotes (see e.g., Leyh, *et al.*, 1992. *J. Biol. Chem.* **267**: 10405-10410; Schwedock and Long, 1989. *Mol. Plant Microbe Interaction* **2**: 181-194; Laue and Nelson, 1994. *J. Bacteriol.* **176**: 3723-3729); eukaryotes (see e.g., Cherest, *et al.*, 1987. *Mol. Gen. Genet.* **210**: 307-313; Mountain and Korch, 1991. *Yeast* **7**: 873-880; Foster, *et al.*, 1994. *J. Biol. Chem.* **269**: 19777-19786); plants (see e.g., Leustek, *et al.*, 1994. *Plant Physiol.* **105**: 897-90216); and animals (see e.g., Li, *et al.*, 1995. *J. Biol. Chem.* **270**:

29453-29459). The enzyme is homo-oligomer or heterodimer, depending upon the specific source (see e.g., Leyh and Suo, 1992. *J. Biol. Chem.* **267**: 542-545).

ATP sulfurylase has been used for many different applications, for example, bioluminometric detection of ADP at high concentrations of ATP (see e.g., Schultz, *et al.*, 1993. *Anal. Biochem.* **215**: 302-304); continuous monitoring of DNA polymerase activity (see e.g., Nyrbn, 1987. *Anal. Biochem.* **167**: 235-238); and DNA sequencing (see e.g., Ronaghi, *et al.*, 1996. *Anal. Biochem.* **242**: 84-89; Ronaghi, *et al.*, 1998. *Science* **281**: 363-365; Ronaghi, *et al.*, 1998. *Anal. Biochem.* **267**: 65-71).

Several assays have been developed for detection of the forward ATP sulfurylase reaction. The colorimetric molybdolysis assay is based on phosphate detection (see e.g., Wilson and Bandurski, 1958. *J. Biol. Chem.* **233**: 975-981), whereas the continuous spectrophotometric molybdolysis assay is based upon the detection of NADH oxidation (see e.g., Seubert, *et al.*, 1983. *Arch. Biochem. Biophys.* **225**: 679-691; Seubert, *et al.*, 1985. *Arch. Biochem. Biophys.* **240**: 509-523). The later assay requires the presence of several detection enzymes. In addition, several radioactive assays have also been described in the literature (see e.g., Daley, *et al.*, 1986. *Anal. Biochem.* **157**: 385-395). For example, one assay is based upon the detection of $^{32}\text{PP}_i$ released from ^{32}P -labeled ATP (see e.g., Seubert, *et al.*, 1985. *Arch. Biochem. Biophys.* **240**: 509-523) and another on the incorporation of ^{35}S into $[^{35}\text{S}]$ -labeled APS (this assay also requires purified APS kinase as a coupling enzyme; see e.g., Seubert, *et al.*, 1983. *Arch. Biochem. Biophys.* **225**: 679-691); and a third reaction depends upon the release of $^{35}\text{SO}_4^{-2}$ from $[^{35}\text{S}]$ -labeled APS (see e.g., Daley, *et al.*, 1986. *Anal. Biochem.* **157**: 385-395).

For detection of the reversed ATP sulfurylase reaction a continuous spectrophotometric assay (see e.g., Segel, *et al.*, 1987. *Methods Enzymol.* **143**: 334-349); a bioluminometric assay (see e.g., Balharry and Nicholas, 1971. *Anal. Biochem.* **40**: 1-17); an $^{35}\text{SO}_4^{-2}$ release assay (see e.g., Seubert, *et al.*, 1985. *Arch. Biochem. Biophys.* **240**: 509-523); and a $^{32}\text{PP}_i$ incorporation assay (see e.g., Osslund, *et al.*, 1982. *Plant Physiol.* **70**: 39-45) have been previously described.

ATP produced by an ATP sulfurylase can be converted using enzymatic reactions which convert ATP to light. Light-emitting chemical reactions (*i.e.*, chemiluminescence) and biological reactions (*i.e.*, bioluminescence) are widely used in analytical biochemistry for sensitive

measurements of various metabolites. In bioluminescent reactions, the chemical reaction that leads to the emission of light is enzyme-catalyzed. For example, the luciferin-luciferase system allows for specific assay of ATP and the bacterial luciferase-oxidoreductase system can be used for monitoring of NAD(P)H. Both systems have been extended to the analysis of numerous substances by means of coupled reactions involving the production or utilization of ATP or NAD(P)H (see e.g., Kricka, 1991. *Chemiluminescent and bioluminescent techniques. Clin. Chem.* 37: 1472-1281).

The development of new reagents have made it possible to obtain stable light emission proportional to the concentrations of ATP (see e.g., Lundin, 1982. Applications of firefly luciferase In; *Luminescent Assays* (Raven Press, New York) or NAD(P)H (see e.g., Lovgren, et al., Continuous monitoring of NADH-converting reactions by bacterial luminescence. *J. Appl. Biochem.* 4: 103-111). With such stable light emission reagent, it is possible to make endpoint assays and to calibrate each individual assay by addition of a known amount of ATP or NAD(P)H. In addition, a stable light-emitting system also allows continuous monitoring of ATP- or NAD(P)H-converting systems.

Suitable enzymes for converting ATP into light include luciferases, e.g., insect luciferases. Luciferases produce light as an end-product of catalysis. The best known light-emitting enzyme is that of the firefly, *Photinus pyralis* (*Coleoptera*). The corresponding gene has been cloned and expressed in bacteria (see e.g., de Wet, et al., 1985. *Proc. Natl. Acad. Sci. USA* 80: 7870-7873) and plants (see e.g., Ow, et al., 1986. *Science* 234: 856-859), as well as in insect (see e.g., Jha, et al., 1990. *FEBS Lett.* 274: 24-26) and mammalian cells (see e.g., de Wet, et al., 1987. *Mol. Cell. Biol.* 7: 725-7373; Keller, et al., 1987. *Proc. Natl. Acad. Sci. USA* 82: 3264-3268). In addition, a number of luciferase genes from the Jamaican click beetle, *Pyroplorus plagiophihalamus* (*Coleoptera*), have recently been cloned and partially characterized (see e.g., Wood, et al., 1989. *J. Biolumin. Chemilumin.* 4: 289-301; Wood, et al., 1989. *Science* 244: 700-702). Distinct luciferases can sometimes produce light of different wavelengths, which may enable simultaneous monitoring of light emissions at different wavelengths. Accordingly, these aforementioned characteristics are unique, and add new dimensions with respect to the utilization of current reporter systems.

Firefly luciferase catalyzes bioluminescence in the presence of luciferin, adenosine 5'-triphosphate (ATP), magnesium ions, and oxygen, resulting in a quantum yield of 0.88 (see e.g., McElroy and Selinger, 1960. *Arch. Biochem. Biophys.* **88**: 136-145). The firefly luciferase bioluminescent reaction can be utilized as an assay for the detection of ATP with a 5 detection limit of approximately 1×10^{-13} M (see e.g., Leach, 1981. *J. Appl. Biochem.* **3**: 473-517). In addition, the overall degree of sensitivity and convenience of the luciferase-mediated detection systems have created considerable interest in the development of firefly luciferase-based biosensors (see e.g., Green and Kricka, 1984. *Talanta* **31**: 173-176; Blum, et al., 1989. *J. Biolumin. Chemilumin.* **4**: 543-550).

10 Using the above-described enzymes, the sequence primer is exposed to a polymerase and a known dNTP. If the dNTP is incorporated onto the 3' end of the primer sequence, the dNTP is cleaved and a PPi molecule is liberated. The PPi is then converted to ATP with ATP sulfurylase. Preferably, the ATP sulfurylase is present at a sufficiently high concentration that the conversion of PPi proceeds with first-order kinetics. In the presence of luciferase, the ATP is 15 hydrolyzed to liberate a photon. The reaction preferably has a sufficient concentration of luciferase present within the reaction mixture such that the reaction, $\text{ATP} \rightarrow \text{ADP} + \text{PO}_4^{3-} +$ photon (light), proceeds with first-order kinetics. The photon can be measured using methods and apparatuses described below.

**can be detected by examining a relative increase in sequencing by

20 For most applications it is desirable to wash away diffusible sequencing reagents, e.g., unincorporated dNTPs, with a wash buffer. Any wash buffer used in pyrophosphate sequencing can be used. An example of a wash buffer is 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl, 1% Tween 20 (Nyren et al., *Anal. Biochem.* **208**: 171-75, 1993).

In some embodiments, the concentration of reactants in the sequencing reaction include 1 25 pmol DNA, 3 pmol polymerase, 40 pmol dNTP in 0.2 ml buffer. See Ronaghi, et al., *Anal. Biochem.* **242**: 84-89 (1996).

The sequencing reaction can be performed with each of four predetermined nucleotides, if desired. A "complete" cycle generally includes sequentially administering sequencing 30 reagents for each of the nucleotides dATP, dGTP, dCTP and dTTP (or dUTP), in a predetermined order. Unincorporated dNTPs are washed away between each of the nucleotide

additions. Alternatively, unincorporated dNTPs are degraded by apyrase (see below). The cycle is repeated as desired until the desired amount of sequence of the sequence product is obtained. In some embodiments, about 10-1000, 10-100, 10-75, 20-50, or about 30 nucleotides of sequence information is obtained from one annealed primer.

5 Luciferase can hydrolyze dATP directly with concomitant release of a photon. This results in a false positive signal because the hydrolysis occurs independent of incorporation of the dATP into the extended sequencing primer. To avoid this problem, a dATP analog can be used which is incorporated into DNA, *i.e.*, it is a substrate for a DNA polymerase, but is not a substrate for luciferase. One such analog is α -thio-ATP such as adenosine 5'-phosphosulfate
10 (APS). Thus, use of APS avoids the spurious photon generation that can occur when dATP is hydrolyzed without being incorporated into a growing nucleic acid chain.

Typically, the PPi-based detection is calibrated by the measurement of the light released following the addition of control nucleotides to the sequencing reaction mixture immediately after the addition of the sequencing primer. This allows for normalization of the reaction
15 conditions. Incorporation of two or more identical nucleotides in succession is revealed by a corresponding increase in the amount of light released. Thus, a two-fold increase in released light relative to control nucleotides reveals the incorporation of two successive dNTPs into the extended primer.

If desired, apyrase may be "washed" or "flowed" over the surface of the solid support so
20 as to facilitate the degradation of any remaining, non-incorporated dNTPs within the sequencing reaction mixture. Upon treatment with apyrase, any remaining reactants are washed away in preparation for the following dNTP incubation and photon detection steps. Alternatively, the apyrase may be bound to the solid support.

When the support is planar, the pyrophosphate sequencing reactions preferably take place
25 in a thin, aqueous reaction chamber comprising an optically-transparent solid support surface and an optically-transparent cover. Sequencing reagents may then be delivered by flowing them across the surface of the substrate. When the support is not planar, the reagents may be delivered by dipping the solid support into baths of any given reagents.

In some embodiments, the enzymes utilized in the pyrophosphate sequencing reaction (e.g., sulfurylase, luciferase, and apyrase) may be immobilized onto the solid support. When luciferase is immobilized, it is preferably less than 50 µm from an anchored primer.

The photons generated by luciferase may be quantitated using a variety of detection apparatuses, e.g., a photomultiplier tube, charge-coupled display (CCD), absorbance photometer, and a luminometer, as well as the apparatuses described herein. In a preferred embodiment, the quantitation of the emitted photons is accomplished by the use of charge-coupled display (CCD) camera fitted with a microchannel plate intensifier. CCD detectors are described in, e.g., Bronks, *et al.*, 1995. *Anal. Chem.* 65: 2750-2757. Preferably, the CCD camera uses a custom designed and fabricated CCD possessing a total of 16 million pixels (i.e., 4,000 x 4,000 pixel array) which can detect approximately 1% of the photons produced and can convert 40% to 80% of the photons produced into an actual measurable signal. With this system, approximately 1% of the photons produced are detected. This system can convert 40% to 80% of the photons produced into an actual measurable signal. Additionally, this CCD system possesses a minimum signal-to-noise ratio of 5:1, with a 10:1 signal-to-noise ratio being preferable.

Apparatuses for Sequencing Nucleic Acids

Also provided in the invention are apparatuses for sequencing nucleic acids. In some embodiments, the apparatuses include anchor primers attached to planar substrates. Nucleic acid sequence information can be detected using conventional optics or fiber-optic based systems attached to the planar substrate. In other embodiments, the apparatuses include anchor primers attached to the termini of fiber-optic arrays. In these embodiments, sequence information can be obtained directly from the termini of the fiber optic array.

25

Apparatus for sequencing nucleic acids

An apparatus for sequencing nucleic acids is illustrated in FIG. 2. The apparatus includes an inlet conduit 200 in communication with a detachable perfusion chamber 220. The inlet conduit 200 allows for entry of sequencing reagents via a plurality of tubes 202-212, which are each in communication with a plurality of sequencing dispensing reagent vessels 214-224.

Reagents are introduced through the conduit 200 into the perfusion chamber 220 using either a pressurized system or pumps to drive positive flow. Typically, the reagent flow rates are from 1 to 50 ml/minute with volumes from 0.100 ml to continuous flow (for washing). Valves are under computer control to allow cycling of nucleotides and wash reagents.

5 Sequencing reagents, e.g., polymerase can be either pre-mixed with nucleotides or added in stream. A manifold brings all six tubes 202-212 together into one for feeding the perfusion chamber. Thus several reagent delivery ports allow access to the perfusion chamber. For example, one of the ports may be utilized to allow the input of the aqueous sequencing reagents, while another port allows these reagents (and any reaction products) to be withdrawn from the

10 perfusion chamber.

The perfusion chamber 200 contains a substrate to which a plurality of anchor primers have been attached. This can be a planar substrate containing one or more anchored primers in anchor pads formed at the termini of a bundled fiber optic arrays. The latter substrate surface is discussed in more detail below.

15 The perfusion chamber allows for a uniform, linear flow of the required sequencing reagents, in aqueous solution form, over the amplified nucleic acids and allows for the rapid and complete exchange of these reagents. Thus, it is suitable for performing pyrophosphate-based sequencing reaction. The perfusion chamber can also be used to prepare the anchor primers and perform amplification reactions, e.g., the RCA reactions described herein.

20 The perfusion chamber is linked to an imaging system 230, which includes a CCD system in association with conventional optics or a fiber optic bundle. For DNA immobilized on an anchor pad of 10 μm in diameter, a 100 μm diameter lens for CCD imaging is preferably placed 1 cm above the slide. For fiber-optic based imaging, it is preferable to incorporate the optical fibers directly into the cover slip.

25 The imaging system 230 is used to collect light from the reactors on the substrate surface. Light can be imaged, for example, onto a CCD using a high sensitivity low noise apparatus known in the art.

The imaging system is linked to a computer control and data collection system 240. In general, any commonly available hardware and software package can be used. The computer 30 control and data collection system is also linked to the conduit 200 to control reagent delivery.

An example of a perfusion chamber of the present invention is illustrated in FIG. 3. The perfusion chamber includes a sealed compartment with transparent upper and lower slide. It is designed to allow linear flow of solution over the surface of the substrate surface and to allow for fast exchange of reagents. Thus, it is suitable for carrying out, for example, the 5 pyrophosphate sequencing reactions. Laminar flow across the perfusion chamber can be optimized by decreasing the width and increasing the length of the chamber.

The perfusion chamber is preferably detached from the imaging system while it is being prepared and only placed on the imaging system when sequencing analyses is performed.

In one embodiment, the solid support (*i.e.*, a DNA chip or glass slide) is held in place by 10 a metal or plastic housing, which may be assembled and disassembled to allow replacement of said solid support.

The lower side of the solid support of the perfusion chamber carries the reaction center array and, with a traditional optical-based focal system, a high numerical aperture objective lens is used to focus the image of the reaction center array onto the CCD imaging system.

15 The photons generated by the pyrophosphate sequencing reaction are captured by the CCD only if they pass through a focusing device (*e.g.*, an optical lens or optical fiber) and are focused upon a CCD element. However, the emitted photons should escape equally in all directions. In order to maximize their subsequent "capture" and quantitation when utilizing a planar array (*e.g.*, a DNA chip), it is preferable to collect the photons immediately at the planar 20 solid support (*e.g.*, the cover slip). This is accomplished by either: (i) utilizing optical immersion oil between the cover slip and a traditional optical lens or optical fiber bundle or, preferably, (ii) incorporating optical fibers directly into the cover slip itself. Similarly, when a thin, optically-transparent planar surface is used, the optical fiber bundle can also be placed against its back surface, eliminating the need to "image" through the depth of the entire reaction/perfusion 25 chamber.

In some embodiments, the solid support is coupled to a bundle of optical fibers, which are used to detect and transmit sequence reaction of byproducts. The total number of optical fibers within the bundle may be varied so as to match the number of individual arrays utilized in the sequencing reaction. The number of optical fibers incorporated into the bundle is designed to 30 match the CCD (*i.e.*, approximately 30 mm x 30mm) so as to allow 1:1 imaging. Commercially-

available optical fiber bundles range from 25 cm x 25 cm to 10 μm in diameter. The desired number of optical fibers are initially fused into a bundle, the terminus of which is cut and polished so as to form a "wafer" of the required thickness (e.g., 1.5 mm). The resulting optical fiber wafers possess similar handling properties to that of a plane of glass. The individual fibers
5 can be any size diameter (e.g., 3 μm to 100 μm).

Fiber optic substrate arrays with linked anchor primers

In other embodiments, the planar support is omitted and the anchor primers are linked directly to the termini of the optical fibers. Preferably, the anchor primers are attached to termini
10 that are cavitated as shown schematically in FIG. 4. The termini are treated, e.g., with acid, to form a hemispherical shape indentation, or cavitation, that ranges from approximately one-half the depth of an individual optical fiber up to 2- to 3-times the diameter of the fiber. When used for pyrophosphate-based sequencing, the cavity is preferably 50 μm deep.

Cavities can be introduced into the termini of the fibers by placing one side of the optical
15 fiber wafer into an acid bath for a variable amount of time. The amount of time can vary depending upon the overall depth of the reaction cavity desired (see e.g., Walt, *et al.*, 1996. *Anal. Chem.* 70: 1888). Several methods are known in the art for attaching molecules (and detecting
the attached molecules) in the cavities etched in the ends of fiber optic bundles. See, e.g.,
Michael, *et al.*, *Anal. Chem.* 70: 1242-1248 (1998); Ferguson, *et al.*, *Nature Biotechnology* 14:
20 1681-1684 (1996); Healey and Walt, *Anal. Chem.* 69: 2213-2216 (1997). A pattern of reactive sites can also be created in the microwell, using photolithographic techniques similar to those used in the generation of a pattern of reaction pads on a planar support. See, Healey, *et al.*,
Science 269: 1078-1080 (1995); Munkholm and Walt, *Anal. Chem.* 58: 1427-1430 (1986), and
Bronk, *et al.*, *Anal. Chem.* 67: 2750-2757 (1995).

25 The opposing side of the optical fiber wafer (*i.e.*, the non-etched side) is highly polished so as to allow optical-coupling (e.g., by immersion oil or other optical coupling fluids) to a second, optical fiber bundle. This second optical fiber bundle exactly matches the diameter of the optical wafer containing the reaction chambers, and serve to act as a conduit for the transmission of the photons, generated by the pyrophosphate sequencing reaction, to its attached
30 CCD imaging system or camera.

In a preferred embodiment, the individual optical fibers utilized to generate the fused optical fiber bundle/wafer are larger in diameter (*i.e.*, 6 μm to 12 μm) than those utilized in the optical imaging system (*i.e.*, 3 μm). Thus, several of the optical imaging fibers can be utilized to image a single reaction site.

5 The etched, hemispherical geometry allows for simultaneously reducing background signal from the PP_i released from adjacent anchor pads. In contrast to use of a "chip"-based geometry, wherein the required sequencing reagents are "flowed" over the surface of the solid support matrix (*i.e.*, the anchor pads), delivery of the various sequencing reagents in acid-etched optical fiber bundle embodiment is performed by immersion of the acid-etched cavities, 10 alternately, into dNTP/APS/sulfurylase reagents and then, subsequently, into the apyrase reagents to facilitate the degradation of any remaining dNTPs.

15 It has been unexpectedly found that this system is markedly more efficient than the currently-utilized CCD capture techniques. For a hemispherical-shaped acid-etched cavity , approximately 85% of the emitted photons will impinge upon, and be directed down the length of the individual optical fiber to the CCD camera. Thus, fewer numbers of rolling circle amplification reactions are required to generate a detectable signal.

Mathematical analysis underlying optimization of the pyrophosphate sequencing reaction

20 While not wishing to be bound by theory, it is believed that optimization of reaction conditions can be optimized using assumptions underlying the following analyses.

Solid-phase pyrophosphate sequencing was initially developed by combining a solid-phase technology and a sequencing-by-synthesis technique utilizing bioluminescence (see *e.g.*, Ronaghi, *et al.*, 1996. Real-time DNA sequencing using detection of pyrophosphate release. 25 *Anal. Biochem.* **242**: 84-89). In the solid-phase methodology, an immobilized, primed DNA strand is incubated with DNA polymerase, ATP sulfurylase, and luciferase. By stepwise nucleotide addition with intermediate washing, the event of sequential polymerization can be followed. A remarkable increase in signal-to-noise ratio was obtained by the use of α -thio dATP in the system. This dATP analog is demonstrated to be efficiently incorporated by DNA 30 polymerase while being silent for luciferase, allowing the sequencing reaction to be performed in

real-time. In these early studies, sequencing of a PCR product using streptavidin-coated magnetic beads as a solid support was presented. However, it was found that the loss of the beads during washing, which was performed between each nucleotide and enzyme addition, was the limiting factor to sequence longer stretches.

5 Currently, pyrophosphate sequencing methodologies have a reasonably well-established history for ascertaining the DNA sequence from many identical copies of a single DNA sequencing template (see e.g., Ronaghi, *et al.*, 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release, *Anal. Biochem.* **242**: 84-89; Nyrén, *et al.*, Method of Sequencing DNA, patent WO9813523A1 (issued April 2, 1998; filed Sept. 26, 1997); Ronaghi, 10 *et al.*, 1998. A Sequencing Method Based on Real-Time Pyrophosphate *Science* **281**: 363-365 (1998). Pyrophosphate (PPi)-producing reactions can be monitored by a very sensitive technique based on bioluminescence (see e.g., Nyrén, *et al.*, 1996. pp. 466-496 (*Proc. 9th Inter. Symp. Biolumin. Chemilumin.*)). These bioluminometric assays rely upon the detection of the PPi released in the different nucleic acid-modifying reactions. In these assays, the PPi which is 15 generated is subsequently converted to ATP by ATP sulfurylase and the ATP production is continuously monitored by luciferase. For example, in polymerase-mediated reactions, the PPi is generated when a nucleotide is incorporated into a growing nucleic acid chain being synthesized by the polymerase. While generally, a DNA polymerase is utilized to generate PPi during a pyrophosphate sequencing reaction (see e.g., Ronaghi, *et al.*, 1998. *Doctoral Dissertation*, The 20 Royal Institute of Technology, Dept. of Biochemistry (Stockholm, Sweden)), it is also possible to use reverse transcriptase (see e.g., Karamohammed, *et al.*, 1996. pp. 319-329 (*Proc. 9th Inter. Symp. Biolumin. Chemilumin.*)) or RNA polymerase (see e.g., Karamohammed, *et al.*, 1998. *BioTechniques* **24**: 302-306) to follow the polymerization event.

For example, a bioluminometric primer extension assay has been utilized to examine 25 single nucleotide mismatches at the 3'-terminus (see e.g., Nyrén, *et al.*, 1997. *Anal. Biochem.* **244**: 367-373). A phage promoter is typically attached onto at least one of the arbitrary primers and, following amplification, a transcriptional unit may be obtained which can then be subjected to stepwise extension by RNA polymerase. The transcription-mediated PPi-release can then be detected by a bioluminometric assays (e.g., ATP sulfurylase-luciferase). By using this strategy,

it is likely to be possible to sequence double-stranded DNA without any additional specific sequencing primer. In a series of "run-off" assays, the extension by T₇ phage RNA polymerase has been examined and was found to be rather slow (see e.g., Kwok, *et al.*, 1990. *Nucl. Acids Res.* **18**: 999-1005). However, the substitution of an α -thio nucleotide analogs for the subsequent, correct natural deoxynucleotide after the 3'-mismatch termini, the rate of polymerization could be decreased by 5-fold to 13-fold, thus causing a delay in the incorporation of correct nucleotides by the DNA polymerase after the primer comprising a mismatch at the 3'-termini. However, after incorporation of a few bases the rate of DNA synthesis is comparable with the rate observed for a normal template/primer. Single-base detection by this technique has been improved by incorporation of apyrase to the system, which functions to degrade the nucleotide to a concentration far below the K_m of the DNA polymerase. The use of apyrase minimizes further extension upon contact with a mismatched base, and thereby simplifies the data analysis. The above-described technique provides a rapid and real-time analysis for applications in the areas of mutation detection and single-nucleotide polymorphism (SNP) analysis.

The pyrophosphate sequencing system takes advantage of the cooperativity of several enzymes to monitor DNA synthesis. Parameters such as stability, fidelity, specificity, sensitivity, K_M and K_{CAT} are of paramount importance for the optimal performance of the enzymes used in the system. In the pyrophosphate sequencing system, the activity of the detection enzymes (*i.e.*, sulfurylase and luciferase) generally remain constant during the sequencing reaction, and are only very slightly inhibited by high amounts of products (see e.g., Ronaghi, *et al.*, 1998. *Doctoral Dissertation*, The Royal Institute of Technology, Dept. of Biochemistry (Stockholm, Sweden)). Sulfurylase converts PPi to ATP in approximately 2.0 seconds, and the generation of light by luciferase takes place in less than 0.2 seconds. The most critical reactions are the DNA polymerization and the degradation of nucleotides. The value of the enzymes utilized in the pyrophosphate sequencing methodology are listed below:

<u>Enzyme</u>	<u>K_M (μM)</u>	<u>K_{CAT} (S^{-1})</u>
Klenow	0.18 (dTTP)	0.92

T, DNA Polymerase	0.36 (dTTP)	0.52
ATP Sulfurylase	0.56 (APS); 7.0 (PPi)	38
Firefly Luciferase	20 (ATP)	0.015
Apyrase	120 (ATP); 260 (ADP)	500 (ATP)

5 The enzymes involved in these two reactions are obviously competing for the same substrate. Therefore, changes in the kinetics of these enzymes directly influence the performance of the sequencing reaction. At the time of dNTP addition, a nucleotide attaches to a polymerase bound to DNA, and polymerization begins. To obtain a rapid polymerization the nucleotide triphosphate concentration must be above the K_M of the DNA polymerase. Conversely, if the
10 concentration of the nucleotide triphosphates is too high, lower fidelity of the polymerase is frequently observed (see e.g., Cline, *et al.*, 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucl. Acids Res.* **24**: 3546-3551), although, the K_M for the misincorporation rate is much higher than that of the rate for correct incorporation (see e.g., Capson, *et al.*, 1992. Kinetic characterization of the polymerase and exonuclease activity of the
15 gene 43 protein of bacteriophage T4. *Biochemistry* **31**: 10984-10994). Although a very high fidelity can be achieved by using polymerases with inherent exonuclease activity, their use also holds the disadvantage that primer degradation may occur.

Although the exonuclease activity of the Klenow fragment of DNA polymerase I (Klenow) is low, it has been demonstrated that the 3'-terminus of the primer was degraded with
20 longer incubations in the absence of nucleotide triphosphates (see e.g., Ronaghi, *et al.*, 1998. *Doctoral Dissertation*, The Royal Institute of Technology, Dept. of Biochemistry (Stockholm, Sweden)). Even in the absence of exonuclease activity, an induced-fit binding mechanism in the polymerization step provides a very efficient selectivity for the correct dNTP with a net contribution, approaching a fidelity of 1×10^5 to 1×10^6 (see e.g., Wong, *et al.*, 1991. An induced-fit kinetic mechanism for DNA replication fidelity. *Biochemistry* **30**: 526-537). In
25 pyrophosphate sequencing, exonuclease-deficient (exo-) polymerases, such as exo-Klenow or Sequenase®, catalyze the incorporation of a nucleotide only in the presence of a complementary dNTP, confirming a high fidelity of these enzymes even in the absence of proof-reading

exonuclease activity. For most polymerases, the K_M and K_{CAT} for a the incorporation of a single nucleotide is lower than that of the incorporation of several (see e.g., Van Draanen, *et al.*, 1992. Beta-L-thymidine 5'-triphosphate analogs as DNA polymerase substrates. *J. Biol. Chem.* **267**: 25019-25024). However, the K_M values for nucleotides are much lower for DNA polymerases, 5 than for apyrase. An increased fidelity in the system can thus be obtained because the nucleotide concentration necessary for efficient polymerization is relatively low and apyrase degrades nucleotides to a concentration far below the K_M of the polymerase in less than 10-15 seconds. The nucleotide-degrading enzyme must possess the following properties: firstly, the enzyme must hydrolyze all deoxynucleotide triphosphates at approximately the same rate. Secondly, it 10 should also hydrolyze ATP to prevent the accumulation of ATP between cycles. Thirdly, the time for nucleotide degradation by the nucleotide-degrading enzyme must be lower than nucleotide incorporation by the polymerase. It is also important that the yield of primer-directed incorporation is as close to 100% as possible before the nucleotide-degrading enzyme has degraded the nucleotide to a concentration below the K_M of the polymerase. Changes in other 15 parameters, such as pH, temperature, and ionic concentration may also alter the kinetics of the different enzymes in the system. However, the enzymes typically utilized in the pyrophosphate sequencing system show high stability within a rather broad range of these parameters for several hours (see e.g., Ronaghi, *et al.*, 1998. *Doctoral Dissertation*, The Royal Institute of Technology, Dept. of Biochemistry (Stockholm, Sweden)).

20 Due to the fact that methodologies currently exist which allow the spectroscopic-detection of single molecules, traditional, the cloning of nucleic acid samples is no longer an absolute requirement in order to obtain nucleic acid sequence information. Currently, a single copy of template which is amplified (e.g., rolling circle amplification) provides a sufficient sample size for the nucleic acid sequencing methodology of the present invention. In brief, the 25 apparatus and methods of the present invention allow the "capture" and quantitation of signals (*i.e.*, photons) within a given optical plane and their subsequent conversion into digital information. Photons are collected from a thin plane roughly equivalent to the volume within which the enzyme and newly synthesized base reside.

Estimates for the spatial and temporal constraints on the pyrophosphate sequencing methodology of the present invention have been calculated, wherein the instant system possesses a 1 cm² area with height approximately 50 µm, for a total volume of 5 µl. With respect to temporal constraints, the molecular species participating in the cascade of reactions are initially defined, wherein:

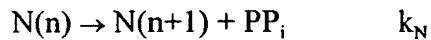
N = the DNA attached to the surface

PP_i = the pyrophosphate molecule released

ATP = the ATP generated from the pyrophosphate

L = the light released by luciferase

It is further specified that N(0) is the DNA with no nucleotides added, N(1) has 1 nucleotide added, N(2) has 2 nucleotides added, and so on. The pseudo-first-order rate constants which relate the concentrations of molecular species are:



In addition, the diffusion constants D_P for PP_i and D_A for ATP must also be specified. These values may be estimated from the following exemplar diffusion constants for biomolecules in a dilute water solution (see Weisiger, 1997. Impact of Extracellular and Intracellular Diffusion on Hepatic Uptake Kinetics Department of Medicine and the Liver Center, University of California, San Francisco, California, USA, dickw@itsa.ucsf.edu, <http://dickw.ucsf.edu/papers/goresky97/chapter.html>).

<u>Molecule</u>	<u>D/10⁻⁵ cm²/sec</u>	<u>Method</u>	<u>Original Reference</u>
Albumin	0.066	lag time	1
Albumin	0.088	light scattering	2
Water	1.940	NMR	3

5 wherein, Original Reference 1 is: Longsworth, 1954. Temperature dependence of diffusion in aqueous solutions, *J. Phys. Chem.* **58**: 770-773; Original Reference 2 is: Gaigalas, *et al.*, 1992. Diffusion of bovine serum albumin in aqueous solutions, *J. Phys. Chem.* **96**: 2355-2359; and Original Reference 3 is: Cheng, 1993. Quantitation of non-Einstein diffusion behavior of water in biological tissues by proton NMR diffusion imaging: Synthetic image calculations,
10 *Magnet. Reson. Imaging* **11**: 569-583.

In order to estimate the diffusion constant of PP_i, the following exemplar values may be utilized (see *CRC Handbook of Chemistry and Physics*, 1983. (W.E. Weast. Ed.) CRC Press, Inc., Boca Raton, FL):

<u>Molecule</u>	<u>D/10⁻⁵ cm²/sec</u>	<u>Molecular Weight/amu</u>
sucrose	0.5226	342.30
mannitol	0.682	182.18
penta-erythritol	0.761	136.15
glycolamide	1.142	N/A
glycine	1.064	75.07

20 The molecular weight of PP_i is 174 amu. Based upon the aforementioned exemplar values, a diffusion constant of approximately 0.7x10⁻⁵ cm²/sec for PP_i is expected.

It should also be noted that the enzymes catalyzing the three pyrophosphate sequencing reactions are thought to follow Michaelis-Menten kinetics (see *e.g.* Stryer, 1988. *Biochemistry*, W. H. Freeman and Company, New York), which may be described:

25 $K_M = [E][S]/[ES]$,

$$\text{velocity} = V_{\max} [S] / (K_M + [S]),$$

$$V_{\max} = k_{\text{turnover}} [E_T]$$

where $[S]$ is the concentration of substrate, $[E]$ is the concentration of free enzyme, $[ES]$ is the concentration of the enzyme-substrate complex, and $[E_T]$ is the total concentration of enzyme = $[E] + [ES]$.

It is preferable that the reaction times are at least as fast as the solution-phase pyrophosphate-based sequencing described in the literature. That rate that a substrate is converted into product is

$$-d[S]/dt = K_{\text{turnover}} [E_T][S]/(K_M + [S])$$

10 The effective concentration of substrate may be estimated from the size of a replicated DNA molecule, at most $(10 \mu\text{m})^3$ and the number of copies (approximately 10,000), yielding a concentration of approximately 17 nm. This is smaller than the K_M for the enzymes described previously, and therefore the rate can be estimated to be

$$-d[S]/dt = (K_{\text{turnover}}/K_M)[E_T][S].$$

15 Thus, with pseudo first-order kinetics, the rate constant for disappearance of substrate depends on K_{turnover} and K_M , which are constants for a given enzyme, and $[E_T]$. Using the same enzyme concentrations reported in the literature will therefore produce similar rates.

The first step in the pyrophosphate sequencing reaction (*i.e.*, incorporation of a new nucleotide and release of PP_i) will now be examined in detail. The preferred reaction conditions 20 are: 1 pmol DNA, 3 pmol polymerase, 40 pmol dNTP in 0.2 ml buffer. Under the aforementioned, preferred reaction conditions, the K_M for nucleotide incorporation for the Klenow fragment of DNA polymerase I is 0.2 μM and for Sequenase 2.0™ (Promega Biotech, Madison, WI) is 0.4 μM , and complete incorporation of 1 base is less than 0.2 sec (see *e.g.*, Ronaghi, *et al.*, 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release, 25 *Anal. Biochem.* **242**: 84-89) with a polymerase concentration of 1.5 nM.

In the preferred 5 μl reaction volume, there are a total of 10,000 anchor primers with 10,000 sequencing primer sites each, or 1×10^8 total extension sites = 0.17 fmol. Results which

have been previously published in the literature suggest that polymerase should be present at 3-times abundance, or a 0.5 fmol, within the reaction mixture. The final concentration of polymerase is then 0.1 nM. It should be noted that these reaction conditions are readily obtained in the practice of the present invention.

5 As previously stated, the time required for the nucleotide addition reaction is no greater than 0.2 sec per nucleotide. Hence, if the reaction is allowed to proceed for a total of T seconds, then nucleotide addition should be sufficiently rapid that stretches of up to $(T/0.2)$ identical nucleotides should be completely filled-in by the action of the polymerase. As will be discussed infra, the rate-limiting step of the pyrophosphate sequencing reaction is the sulfurylase reaction, 10 which requires a total of approximately 2 sec to complete. Accordingly, a total reaction time which allows completion of the sulfurylase reaction, should be sufficient to allow the polymerase to "fill-in" stretches of up to 10 identical nucleotides. In random DNA species, regions of 10 or more identical nucleotides have been demonstrated to occur with a per-nucleotide probability of approximately 4^{-10} , which is approximately 1×10^{-6} . In the 10,000 sequences which are extended 15 from anchor primers in a preferred embodiment of the present invention, each of which will be extended at least 30 nt. and preferably 100 nt., it is expected that approximately one run of 10 identical nucleotides will be present. Thus, it may be concluded that runs of identical nucleotides should not pose a difficulty in the practice of the present invention.

The overall size of the resulting DNA molecule is, preferably, smaller than the size of the 20 anchoring pads (*i.e.*, 10 μ m) and must be smaller than the distance between the individual anchoring pads (*i.e.*, 100 μ m). The radius of gyration of a single-stranded DNA concatemer with N total nucleotides may be mathematically-estimated by the following equation: radius = b $(N/N_0)^{0.6}$, where b is the persistence length and N_0 is the number of nucleotides per persistence 25 length; the exponent 0.6 is characteristic of a self-avoiding walk (see *e.g.*, Doi, 1986. *The Theory of Polymer Dynamics* (Clarendon Press, New York); Flory, 1953. *Principles of Polymer Chemistry* (Cornell University Press, New York)). Using single-stranded DNA as an example, b is 4 nm and N_0 is 13.6 nt. (see *e.g.*, Grosberg, 1994. *Statistical Physics of Macromolecules* (AIP Press, New York)). Using 10,000 copies of a 100-mer, $N = 1 \times 10^6$ and the radius of gyration is 3.3 μ m.

The sulfurylase reaction will now be discussed in detail. The time for the production of ATP from adenosine 5'-phosphosulfate (APS) and PP_i has been estimated to be less than 2 sec (see e.g., Nyrén and Lundin, 1985. *Anal. Biochem.* **151**: 504-509. The reported reaction conditions for 1 pmol PP_i in 0.2 ml buffer (5 nM) are 0.3 U/ml ATP sulfurylase (ATP:sulfate 5 adenyllyltransferase; Prod. No. A8957; Sigma Chemical Co., St. Louis, MO) and 5 μM APS (see e.g., Ronaghi, *et al.*, 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release, *Anal. Biochem.* **242**: 84-89). The manufacturer's information (Sigma Chemical Co., St. Louis, MO) for sulfurylase (470 kD) reports an activity of 5-20 units per mg protein (*i.e.*, one unit will produce 1.0 μmole of ATP from APS and PP_i per minute at pH 8.0 at 30 C), whereas 10 the specific activity has been reported elsewhere as 140 units per mg (see Karamohamed, *et al.*, 1999. Purification, and Luminometric Analysis of Recombinant *Saccharomyces cerevisiae* MET3 Adenosine Triphosphate Sulfurylase Expressed in *Escherichia coli*, *Prot. Express. Purification* **15**: 381-388). Due to the fact that the reaction conditions utilized in the practice of the present invention are similar to those reaction conditions reported in the aforementioned reference, the sulfurylase concentration within the assay was estimated as 4.6 nM. Thus, at the half-maximal rate, [APS] = 0.5 μM and [PP_i] = 7 μM.

In the reaction conditions utilized in the present invention, [PP_i] is approximately 0.17 fmol in 5 μl, or 0.03 nM. The fraction of PP_i which is bound to the enzyme is [E]/K_M, where [E] is the concentration of free enzyme. Since the enzyme concentration is much larger than the PP_i 20 concentration, the total enzyme concentration alone, may be used in the calculations. The fraction of PP_i bound to enzyme is found to be 4.6 nM / 7 μM = 7x10⁻⁴. Therefore, it may be concluded that the PP_i spends most of its time freely diffusing before being converted to ATP.

The mean time for each phosphate (P) to react is $1/k_p = 2$ seconds. The root mean square (RMS) distance it diffuses in each direction is approximately $2D_p/k_p$, or $2.8 \times 10^3 \mu\text{m}^2$. The RMS 25 distance in each direction is 53 μm. This value indicates that each of the individual anchor primers must be more than 50 μm apart, or PP_i which is released from one anchor could diffuse to the next, and be detected.

Another method which may be used to explain the aforementioned phenomenon is to estimate the amount of PP_i over a first anchor pad that was generated at said first anchor pad

relative to the amount of PP_i that was generated at a second anchor pad and subsequently diffused over to the location of said first anchor pad. When these two quantities approach each other in magnitude, it becomes difficult to distinguish the "true" signal from that of the background. This may be mathematically-described by defining a as the radius of an anchor pad
5 and $1/b^2$ as the density of an anchor pad. Based upon previously published data, a is approximately equal to 10 μm and b is approximately equal to 100 μm . The amount of PP_i, which is present over said first anchor pad may be described by: $\exp(-k_p t)[1 - \exp(-a^2/2D_p t)]$ and the amount of PP_i present over the second anchor pads may be mathematically-approximated by:
10 $(1/3)\exp(-k_p t)[pa^2/b^2]\exp(-b^2/2D_p t)$. The prefactor 1/3 assumes that $1/4$ of the DNA sequences will incorporate 1 nucleotide, $1/4$ of these will then incorporate a second nucleotide, etc., and thus the sum of the series is 1/3. The amounts of PP_i over the first and second anchor pads become similar in magnitude when $2D_p t$ is approximately equal to b^2 , thus indicating that the RMS distance a molecule diffuses is equal to the distance between adjacent anchor pads. In accord,
15 based upon the assay conditions utilized in the practice of the present invention, the anchor pads must be placed no closer than approximately 50 μm apart, and preferable are at least 3-times further apart (*i.e.*, 150 μm).
15

Although the aforementioned findings set a limit on the surface density of anchor pads, it is possible to decrease the distance requirements, while concomitantly increasing the overall surface density of the anchor pads, by the use of a number of different approaches. One
20 approach is to detect only the early light, although this has the disadvantage of losing signal, particularly from DNA sequences which possess a number of contiguous, identical nucleotides.
20

A second approach to decrease the distance between anchor pads is to increase the concentration of sulfurylase in the reaction mixture. The reaction rate k_p is directly proportional to the sulfurylase concentration, and the diffusion distance scales as $k_p^{-1/2}$. Therefore, if the
25 sulfurylase enzyme concentration is increased by a factor of 4-times, the distance between individual anchor pads may be concomitantly reduced by a factor of 2-times.
25

A third approach is to increase the effective concentration of sulfurylase (which will also work for other enzymes described herein) by binding the enzyme to the surface of the anchor pads. The anchor pad can be approximated as one wall of a cubic surface enclosing a sequencing
25

reaction center. Assuming a $10 \mu\text{m} \times 10 \mu\text{m}$ surface for the pad, the number of molecules bound to the pad to produce a concentration of a $1 \mu\text{M}$ is approximately 600,000 molecules.

The sulfurylase concentration in the assay is estimated as 5nM. The number bound molecules to reach this effective concentration is about 3000 molecules. Thus, by binding more 5 enzyme molecules, a greater effective concentration will be attained. For example, 10,000 molecules could be bound per anchor pad.

As previously estimated, each sulfurylase molecule occupies a total area of 65 nm^2 on a surface. Accordingly, anchoring a total of 10,000 sulfurylase enzyme molecules on a surface (*i.e.*, so as to equal the 10,000 PP_i released) would require $1.7 \mu\text{m}^2$. This value is only 10 approximately 2% of the available surface area on a $10 \mu\text{m} \times 10 \mu\text{m}$ anchor pad. Hence, the concentration of the enzyme may be readily increased to a much higher value.

A fourth approach to allow a decrease in the distance between individual anchor pads, is to utilize one or more agents to increase the viscosity of the aqueous-based, pyrophosphate sequencing reagents (*e.g.*, glycerol, polyethylene glycol (PEG), and the like) so as to markedly 15 increase the time it takes for the PP_i to diffuse. However, these agents will also concomitantly increase the diffusion time for other non-immobilized components within the sequencing reaction, thus slowing the overall reaction kinetics. Additionally, the use of these agents may also function to chemically-interfere with the sequencing reaction itself.

A fifth, and preferred, methodology to allow a decrease in the distance between 20 individual anchor pads, is to conduct the pyrophosphate sequencing reaction in a spatial-geometry which physically-prevents the released PP_i from diffusing laterally. For example, uniform cavities, which are generated by acid-etching the termini of optical fiber bundles, may be utilized to prevent such lateral diffusion of PP_i (see Michael, *et al.*, 1998. Randomly Ordered Addressable High-Density Optical Sensor Arrays, *Anal. Chem.* 70: 1242-1248). In this 25 embodiment, the important variable involves the total diffusion time for the PP_i to exit a cavity of height **h**, wherein **h** is the depth of the etched cavity. This diffusion time may be calculated utilizing the equation: $2D_p t = h^2$. By use of the preferred pyrophosphate sequencing reaction conditions of the present invention in the aforementioned calculations, it may be demonstrated that a cavity 50 μm in depth would be required for the sequencing reaction to proceed to

completion before complete diffusion of the PP_i from said cavity. Moreover, this type of geometry has the additional advantage of concomitantly reducing background signal from the PP_i released from adjacent anchor pads. In contrast to use of a "chip"-based geometry, wherein the required sequencing reagents are "flowed" over the surface of the solid support matrix (*i.e.*, the anchor pads), delivery of the various sequencing reagents in acid-etched optical fiber bundle embodiment is performed by immersion of the acid-etched cavities, alternately, into dNTP/APS/sulfurylase reagents and then, subsequently, into the apyrase reagents to facilitate the degradation of any remaining dNTPs.

Subsequently, once ATP has been formed by use of the preferred reaction conditions of the present invention, the reaction time, 1/k_A, has been shown to be 0.2 seconds. Because this reaction time is much lower than the time which the PP_i is free to diffuse, it does not significantly alter any of the aforementioned conclusions regarding the assay geometry and conditions utilized in the present invention.

In order to mitigate the generation of background light, it is preferable to "localize" (*i.e.*, anchoring or binding) the luciferase in the region of the DNA sequencing templates. It is most preferable to localize the luciferase to a region that is delineated by the distance a PP_i molecule can diffuse before it forms ATP. Methods for binding luciferase to a solid support matrix are well-known in the literature (see *e.g.*, Wang, *et al.*, 1997. Specific Immobilization of Firefly Luciferase through a Biotin Carboxyl Carrier Protein Domain, *Analytical Biochem.* 246: 133-139). Thus, for a 2 second diffusion time, the luciferase is anchored within a 50 μm distance of the DNA strand. It should be noted, however, that it would be preferable to decrease the diffusion time and thus to further limit the surface area which is required for luciferase binding.

In order to determine the concentration of luciferase which it is necessary to bind, previously published conditions were utilized in which luciferase is used at a concentration which gives a response of 200 mV for 0.1 μM ATP (see Ronaghi, *et al.*, 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release, *Analytical Biochem.* 242: 84-89). More specifically, it is known from the literature that, in a 0.2 ml reaction volume, 2 ng of luciferase gives a response of 10 mV for 0.1 μM ATP (see Karamohamed and Nyrén, 1999. Real-Time Detection and Quantification of Adenosine Triphosphate Sulfurylase Activity by a

Bioluminometric Approach, *Analytical Biochem.* 271: 81-85). Accordingly, a concentration of 20 ng of luciferase within a 0.2 ml total reaction volume would be required to reproduce these previously-published literature conditions. In the volume of a 10 μ m cube around each of the individual anchor pads of the present invention, a luciferase concentration of 1x10⁻¹⁶ grams
5 would be required, and based upon the 71 kD molecular weight of luciferase, this concentration would be equivalent to approximately 1000 luciferase molecules. As previously stated, the surface area of luciferase has been computed at 50 nm². Thus, assuming the luciferase molecules were biotinylated and bound to the anchor pad, 1000 molecules would occupy a total area of 0.05
10 μ m². From these calculations it becomes readily apparent that a plethora of luciferase molecules may be bound to the anchor pad, as the area of each anchor pad area is 100 μ m².

Again, based upon previously-published results in the literature, each nucleotide takes approximately 3 seconds in toto, to sequence (*i.e.*, 0.5 seconds to add a nucleotide; 2 seconds to make ATP; 0.2 seconds to get fluorescence). Accordingly, a cycle time of approximately 60 seconds per nucleotide is reasonable, requiring approximately 30 minutes per experiment to
15 generate 30 nucleotides of information per sequencing template.

In an alternative embodiment to the aforementioned sequencing methodology (*i.e.*, polymerase → PP_i → sulfurylase → ATP → luciferase → light cascade), a polymerase may be developed (*e.g.*, through the use of protein fusion and the like) which possesses the ability to generate light when it incorporates a nucleotide into a growing DNA chain. In yet another
20 alternative embodiment, a sensor may be developed which directly measures the production of PP_i in the sequencing reaction. As the production of PP_i changes the electric potential of the surrounding buffer, this change could be measured and calibrated to quantitate the concentration of PP_i produced.

As previously discussed, the polymerase-mediated incorporation of dNTPs into the
25 nucleotide sequence in the pyrophosphate sequencing reaction causes the release of an inorganic pyrophosphate (PPi) moiety which, in turn, through catalysis by luciferase, causes the release of a photon (*i.e.*, light). The photons generated by the pyrophosphate sequencing reaction may subsequently be “captured” and quantitated by a variety of methodologies including, but not

limited to: a photomultiplier tube, charge-coupled display (CCD), absorbance photometer, a luminometer, and the like.

The photons generated by the pyrophosphate sequencing reaction are captured by the CCD only if they pass through a focusing device (*e.g.*, an optical lens or optical fiber) and are focused upon a CCD element. The fraction of these photons which are captured may be estimated by the following calculations. First, it is assumed that the lens that focuses the emitted photons is at a distance r from the surface of the solid surface (*i.e.*, DNA chip or etched fiber optic well), where $r = 1$ cm, and that the photons must pass through a region of diameter b (area = $\pi b^2/4$) so as to be focused upon the array element, where $b = 100 \mu\text{m}$. It should also be noted that the emitted photons should escape equally in all directions. At distance r , the photons are dispersed over an area of which is equal to $4\pi r^2$. Thus, the fraction of photons which pass through the lens is described by: $(1/2)[1 - (1 + b^2/4r^2)^{-1/2}]$. When the value of r is much larger than that of b , the fraction which pass through the lens may then be described by: $b^2/16r^2$. For the aforementioned values of r and b , this fraction of photons is 6×10^{-6} .

For each nucleotide addition, it is expected that approximately 10,000 PP_i molecules will be generated and, if all are converted by sulfurylase and luciferase, these PP_i will result in the emission of approximately 1×10^4 photons. In order to maximize their subsequent "capture" and quantitation when utilizing a planar array (*e.g.*, a DNA chip), it is preferable to collect the photons immediately at the planar solid support (*e.g.*, the cover slip). This may be accomplished by either: (*i*) utilizing optical immersion oil between the cover slip and a traditional optical lens or optical fiber bundle or, preferably, (*ii*) incorporating optical fibers directly into the cover slip itself. Performing the previously described calculations (where in this case, $b = 100 \mu\text{m}$ and $r = 50 \mu\text{m}$), the fraction collected is found to be 0.15, which equates to the capture of approximately 1×10^3 photons. This value would be sufficient to provide an adequate signal.

The sequence acquisition software acquires and analyzes the data during the pyrophosphate sequencing cycle. Prior to beginning a given sequencing experiment, a bin of pixels containing each individual reaction center is determined. During each sequencing cycle, four "images" of the entire array are produced, and each image corresponds to excitation of one of the four, fluorescently-labeled nucleotide bases A, C, G, or T (or U). For each reaction center

bin, all of the four images are analyzed to determine which nucleotide species has been incorporated at that reaction center during that cycle. As described above, the reaction center bin corresponding to a certain reaction center contains a 10 x 10 array of pixels. The total number of photons produced by the single fluorophore in that reaction center is determined by the 5 summation of each pixel value in the array. The sums of the reaction center bins from each of the four images are compared, and the image that produces a significant sum corresponds to the newly incorporated base at that reaction center. The images are processed for each of the reaction centers and an array of incorporated nucleotides is recorded. Such processing is capable of being rapidly performed in real-time with modern image processing computers.

10 Multiple “reads” of individual reaction center arrays may be necessary during the detection step to ensure that the four nucleotides are properly distinguished. Exposure times can be as rapid as 100 msec, with the readout time of the CCD chip being on the order of 250 msec. Thus, the maximum time needed for four complete reads of the array is 1.5 seconds. The total 15 time for a given cycle, including reagent addition, removal, and washes, is certainly less than 10 seconds. Accordingly, a sequencing apparatus consisting of an array of 10,000 reaction centers (*i.e.*, a 100 x 100 array) is able to detect at least 360 bases per site per hour, or 3.6 Megabases per hour of total sequence, as a conservative estimate. This rate is significantly faster than those of traditional sequencing methodologies.

20 In addition to short sequencing times, the methods of the present invention do not require the time-consuming processes of initial sample amplification (*e.g.*, cloning or PCR), and gel electrophoresis. The lack of consumables necessary for sample amplification and electrophoresis, coupled with small reagent volumes and reduced manual labor requirements drastically reduce the cost per nucleotide sequenced relative to traditional sequencing techniques.

25 In one embodiment of the present invention, a nucleic acid sample is sheared prior to inclusion in a reaction center. Once these fragments have been sequenced, sequence analysis software is used to assemble their sequences into contiguous stretches. Many algorithms exist in the art that can compare sequences and deduce their correct overlap. New algorithms have recently been designed to process large amounts of sequence data from shotgun (random) sequencing approaches.

In one preferred embodiment, an algorithm initially reduces the amount of data to be processed by using only two smaller sequences derived from either end of the sequence deduced from a single reaction center in a given experiment. This approach has been proposed for use in shotgun sequencing of the human genome (see e.g., Rawlinson, *et al.*, 1996 *J. Virol* 70: 8833-8849; Venter, *et al.*, 1998. *Science* 280: 1540-1542). It employs algorithms developed at the Institute for Genome Research (TIGR; see e.g., Sutton, *et al.*, 1995. *Genome Sci. Technol.* 1: 9-16).

In an alternative, preferred embodiment, raw data is compressed into a fingerprint of smaller words (e.g., hexanucleotide restriction enzyme sites) and these fingerprints can be compared and assembled into larger continuous blocks of sequence (*i.e.*, contigs). This technique is similar to that used to deduce overlapping sequences after oligonucleotide hybridization (see e.g., Idury and Waterman, 1995. *J. Comput. Biol.* 2: 291-306). Yet another embodiment uses existing sequence data, from genetic or physical linkage maps, to assist the assembly of new sequence data from whole genomes or large genomic pieces.

15

The following examples are meant to illustrate, not limit, the invention.

Example 1. Construction of Anchor Primers Linked to a Cavitated Terminus
Fiber Optic Array

20 The termini of a thin wafer fiber optic array are cavitated by inserting the termini into acid as described by Healey *et al.*, *Anal. Chem.* 69: 2213-2216 (1997).

A thin layer of a photoactivatable biotin analog is dried onto the cavitated surface as described Hengsakul and Cass (*Bioconjugate Chem.* 7: 249-254, 1996) and exposed to white light through a mask to create defined pads, or areas of active biotin. Next, avidin is added and 25 allowed to bind to the biotin. Biotinylated oligonucleotides are then added. The avidin has free biotin binding sites that can anchor biotinylated oligonucleotides through a biotin-avidin-biotin link.

The pads are approximately 10 μm on a side with a 100 μm spacing. Oligonucleotides are added so that approximately 37% of the pads include one anchored primer. On a 1 cm surface 30 are deposited 10,000 pads, yielding approximately 3700 pads with a single anchor primer.

Sulfurylase, apyrase, and luciferase are also attached to the cavitated substrate using biotin-avidin.

5 **Example 2. Annealing and Amplification of Members of a Circular Nucleic Acid Library**

A library of open circle library templates is prepared from a population of nucleic acids suspected of containing a single nucleotide polymorphism on a 70 bp Sau3A1-MspI fragment. The templates include adapters that are complementary to the anchor primer, a region 10 complementary to a sequencing primer, and an insert sequence that is to be characterized. The library is generated using Sau3A1 and MspI to digest the genomic DNA. Inserts approximately 65-75 nucleotides are selected and ligated to adaptor oligonucleotides 12 nucleotides in length. The adapter oligonucleotides have sequences complementary to sequences to an anchor primers linked to a substrate surface as described in Example 1.

15 The library is annealed to the array of anchor primers. A DNA polymerase is added, along with dNTPS, and rolling circle replication is used to extend the anchor primer. The result is a single DNA strand, still anchored to the solid support, that is a concatenation of multiple copies of the circular template. 10,000 or more copies of circular templates in the hundred nucleotide size range.

20

Example 3. Sequence Analysis of Nucleic Acid Linked to the Terminus of a Fiber Optic Substrate

25 The fiber optic array wafer containing amplified nucleic acids as described in Example 2 is placed in a perfusion chamber and attached to a bundle of fiber optic arrays, which are themselves linked to a 16 million pixel CCD cameras. A sequencing primer is delivered into the perfusion chamber and allowed to anneal to the amplified sequences.

The sequencing primer primes DNA synthesis extending into the insert suspected of 30 having a polymorphism, as shown in FIG.1. The sequencing primer is first extended by delivering into the perfusion chamber, in succession, a wash solution, a DNA polymerase, and one of dTTP, dGTP, dCTP, or APS (a dATP analog). The sulfurylase, luciferase, and apyrase,

attached to the termini convert any PP_i liberated as part of the sequencing reaction to detectable light. The apyrase present degrades any unreacted dNTP. Light is allowed to collect for 3 seconds by a CCD camera linked to the fiber imaging bundle, after which additional wash solution is added to the perfusion chamber for 10 seconds. The next nucleotide is then added,
5 along with polymerase, thereby repeating the cycle.

During the 10 second wash the collected light image is transferred from the CCD camera to a computer. Light emission is analyzed by the computer and used to determine whether the corresponding dNTP has been incorporated into the extended sequence primer. Addition of dNTPS and pyrophosphate sequencing reagents is repeated until the sequence of the insert region
10 containing the suspected polymorphism is obtained.

Other Embodiments

It is to be understood that while the invention has been described in conjunction with
15 the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

For example, the sequence of the amplified nucleic acid can be determined using by products of RNA synthesis. In this embodiment, an RNA transcript is generated from a
20 promoter sequence present in the circular nucleic acid template library. Suitable promoter sites and their cognate RNA polymerases include RNA polymerases from *E. coli*, the RNA polymerase from the bacteriophage T₃, the RNA polymerase from the bacteriophage T₇, the RNA polymerase from the bacteriophage SP6, and the RNA polymerases from the viral families of bromoviruses, tobamoviruses, tombusvirus, lentiviruses, hepatitis C-like viruses, and
25 picornaviruses. To determine the sequence of an RNA transcript, a predetermined NTP, *i.e.*, an ATP, CTP, GTP, or UTP, is incubated with the template in the presence of the RNA polymerase. Incorporation of the test NTP into a nascent RNA strand can be determined by assaying for the presence of PP_i using the enzymatic detection discussed herein.

WHAT IS CLAIMED IS:

1. A method for sequencing a nucleic acid, the method comprising:
providing one or more or more nucleic acid anchor primers linked to a solid support;
providing a plurality of circular nucleic acid templates;
annealing an effective amount of the nucleic acid anchor primer to at least one of the single-stranded circular templates to yield a primed anchor primer-circular template complex;
combining the primed anchor primer-circular template complex with a polymerase to generate multiple copies of the circular nucleic acid template;
annealing an effective amount of a sequencing primer to the circular nucleic acid template to yield a primed sequencing primer-circular nucleic acid template complex;
extending the sequencing primer with a polymerase and a predetermined nucleotide triphosphate to yield a sequencing product and a sequencing reaction byproduct; and
identifying the sequencing reaction byproduct, thereby determining the sequence of the nucleic acid.
2. The method of claim 1, wherein the circular nucleic acid template is single-stranded DNA.
3. The method of claim 1, wherein the circular nucleic acid template is an open circle nucleic acid.
4. The method of claim 1, wherein the circular nucleic acid template is a closed circle nucleic acid.
5. The method of claim 1, wherein the circular nucleic acid template is genomic DNA.
6. The method of claim 1, wherein the circular nucleic acid template is cDNA.

7. The method of claim 1, wherein the circular nucleic acid is 10-200 nucleotides in length.
8. The method of claim 1, wherein the circular nucleic acid is 10-100 nucleotides in length.
9. The method of claim 1, wherein the circular nucleic acid is 10-50 nucleotides in length.
10. The method of claim 1, wherein the multiple copies are generated by a polymerase chain reaction.
11. The method of claim 1, wherein the primed circular template is extended by rolling circle amplification to yield a single-stranded concatamer of the annealed circular nucleic acid template.
12. The method of claim 11, further comprising:
annealing a reverse primer to the single-stranded concatamer to yield a primed concatamer template, and
combining the primed concatamer template with a polymerase enzyme to generate multiple copies of the concatamer template.
13. The method of claim 1, wherein the sequencing byproduct is pyrophosphate.
14. The method of claim 13, wherein the pyrophosphate is detected by contacting the sequencing byproduct with ATP sulfurylase under conditions sufficient to form ATP.
15. The method of claim 14, wherein the ATP is detected with luciferase.
16. The method of claim 13, further comprising apyrase.

17. The method of claim 13, further comprising washing the sequencing product with a wash buffer.

18. The method of claim 17, wherein the wash buffer includes apyrase.

19. The method of claim 1, wherein the anchor primer sequence includes a biotin group.

20. The method of claim 19, wherein the biotin group on the anchor primer is linked to an avidin group on the solid support.

21. The method of claim 1, wherein the anchor primer is conjugated to a biotin-BSA moiety.

22. The method of claim 21, wherein the biotin-BSA moiety on the anchor primer is linked to an avidin-biotin group on the solid support.

23. The method of claim 21, wherein the biotin-BSA moiety on the anchor primer is linked to a BSA group on the solid support in the presence of silane.

24. The method of claim 1, wherein the solid support includes at least one optical fiber.

25. The method of claim 1, wherein the sequencing primer is extended in the presence of a dATP analog.

26. The method of claim 25, wherein the dATP analog is adenosine 5'-phosphosulfate (APS).

27. The method of claim 1, wherein the solid substrate includes two or more anchoring primers separated by approximately 10 μm to approximately 200 μm .

28. The method of claim 27, wherein the solid substrate includes two or more anchoring primers separated by approximately 50 μm to approximately 150 μm .

29. The method of claim 27, wherein the solid substrate includes two or more anchoring primers separated by approximately 100 μm to approximately 150 μm .

30. The method of claim 1, wherein the solid support matrix comprises of a plurality of anchor pads that are covalently linked to the solid support.

31. The method of claim 30, wherein the surface area of each anchor pad is approximately 10 μm^2 .

32. The method of claim 30, wherein and each pad is separated from one another by a distance ranging from approximately 50 μm to approximately 150 μm .

33. A substrate for analyzing a nucleic acid, the substrate comprising:
a cavitated fiber optic surface; and
a nucleic acid sequence linked to the fiber optic surface.

34. The substrate of claim 33, wherein the substrate comprises a plurality of fiber optic surfaces.

35. The substrate of claim 33, wherein the nucleic acid sequence is an anchor primer.

36. The substrate of claim 33, wherein the fiber optic surface includes two or more anchoring primers separated by approximately 10 μm to approximately 200 μm .

37. The substrate of claim 33, wherein the fiber optic surface includes two or more anchoring primers separated by approximately 100 μm to approximately 150 μm .

38. The substrate of claim 33, wherein the fiber optic surface includes two or more anchoring primers separated by approximately 150 μm .

39. The substrate of claim 33, wherein the fiber optic surface includes two or more anchor pads separated by approximately 100 μm to approximately 150 μm .

40. The substrate of claim 39, wherein the surface area of each pad is approximately 10 μm^2 .

41. A substrate with a cavitated surface comprising 10^3 or more groups of oligonucleotides covalently attached to the surface in discrete known regions, the 10^3 or more groups of oligonucleotides occupying a total area of less than 1 cm^2 on said substrate, said groups of oligonucleotides having different nucleotide sequences.

42. The substrate as recited in claim 41 wherein said substrate comprises 10^4 or more different groups of sequences in discrete known regions.

43. The substrate as recited in claim 1 wherein said substrate comprises 10^5 or more different groups of oligonucleotides with known sequences in discrete known regions.

44. The substrate as recited in claim 1 wherein the groups of oligonucleotides are attached to the surface by a linker.

45. An array of more than 1,000 different groups of oligonucleotide molecules with known sequences covalently coupled to a surface of a cavitated substrate, said groups of oligonucleotide molecules each in discrete known regions and differing from other groups of oligonucleotide molecules in monomer sequence, each of said discrete known regions being an

area of less than about 0.01 cm² and each discrete known region comprising oligonucleotides of known sequence, said different groups occupying a total area of less than 1 cm².

46. The array as recited in claim 45 wherein said area is less than 10,000 microns².

47. The array as recited in claim 46 made by the process of:

exposing a first region of said substrate to light to remove photoremovable group from nucleic acids in said first region, and not exposing a second region of said surface to light;

covalently coupling a first nucleotide to said nucleic acids on said part of said substrate exposed to light, said first nucleotide covalently coupled to said photoremovable group;

exposing a part of said first region of said substrate to light, and not exposing another part of said first region of said substrate to light to remove said photoremovable groups;

covalently coupling a second nucleotide to said part of said first region exposed to light; and

repeating said steps of exposing said substrate to light and covalently coupling nucleotides until said more than 500 different groups of nucleotides are formed on said surface.

48. The array as recited in claim 46 comprising more than 10,000 groups of oligonucleotides of known sequences.

49. An apparatus for analyzing a nucleic acid sequence, the apparatus comprising:
a perfusion chamber, wherein the chamber includes a nucleic acid substrate;
a conduit in communication with the perfusion chamber;
an imaging system in communication with the perfusion chamber; and
a data collection system in communication with the imaging system.

50. The apparatus of claim 49, wherein the substrate is a planar substrate.

51. The apparatus of claim 49, wherein the imaging system is a fiber optic system.

52. The apparatus of claim 49, wherein the substrate comprises

a cavitated fiber optic surface in communication with said imaging system; and
a nucleic acid sequence linked to the fiber optic surface.

53. The apparatus of claim 49, wherein the substrate comprises a plurality of fiber optic surfaces, said fiber optic surfaces being in communication with said imaging system.

54. The apparatus of claim 49, wherein the fiber optic surface includes two or more anchoring primers separated by approximately 100 μm to approximately 150 μm .

55. The apparatus of claim 49, wherein the fiber optic surface includes two or more anchoring primers separated by approximately 150 μm .

56. The apparatus of claim 49, wherein the fiber optic surface includes two or more anchor pads separated by approximately 100 μm to approximately 150 μm .

57. The apparatus of claim 49, wherein the surface area of each pad is approximately 5 μm^2 to approximately 20 μm^2 .

58. The apparatus of claim 49, wherein the surface area of each pad is approximately 10 μm^2 .

59. An apparatus for processing a plurality of analyses, the apparatus comprising:
a flow chamber having disposed therein a substrate comprising a plurality of cavitated
surfaces, said cavitated surfaces having disposed thereon nucleic acid molecules;
fluid means for delivering processing reagents from one or more reservoirs to the flow
chamber so that the analytes anchored to the plurality of microparticles are exposed to the
reagents; and
detection means for detecting a sequence of optical signals from each microparticle of the
plurality, each optical signal of the sequence being indicative of an interaction between a
processing reagent and the analyte anchored thereto, wherein said detection means is in
communication with the cavitated surfaces.

60. The apparatus of claim 59, wherein said detection means further comprises signal
tracking means for correlating said optical signals from each of said microparticles in each of
said digital images to form for each said microparticle of said plurality a sequence of said optical
signals.

61. The apparatus of claim 60, wherein said signal tracking means is a CCD camera.

62. The apparatus of claim 59, wherein said analyte is DNA.

ABSTRACT

Disclosed herein are methods and apparatuses for sequencing a nucleic acid. The method includes annealing a population of circular nucleic acid molecules to a plurality of anchor
5 primers linked to a solid support, and amplifying those members of the population of circular nucleic acid molecules which anneal to the target nucleic acid, and then sequencing the amplified molecules by detecting the presence of a sequence byproduct.

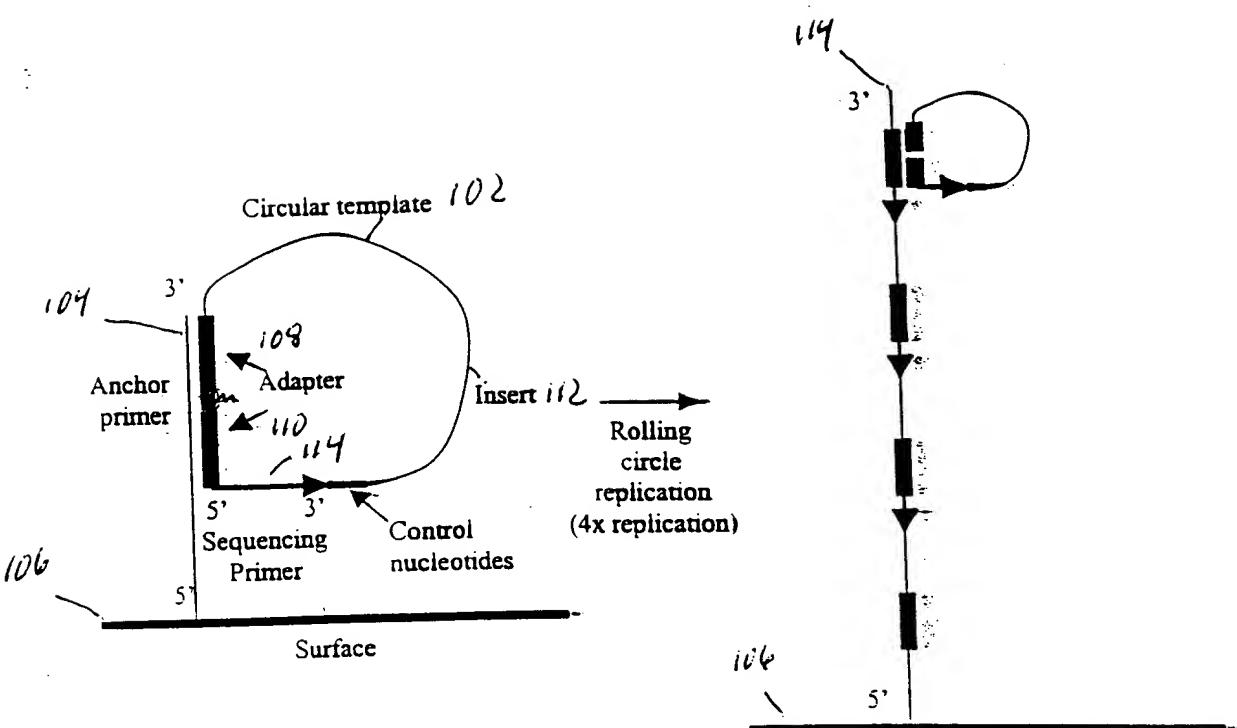
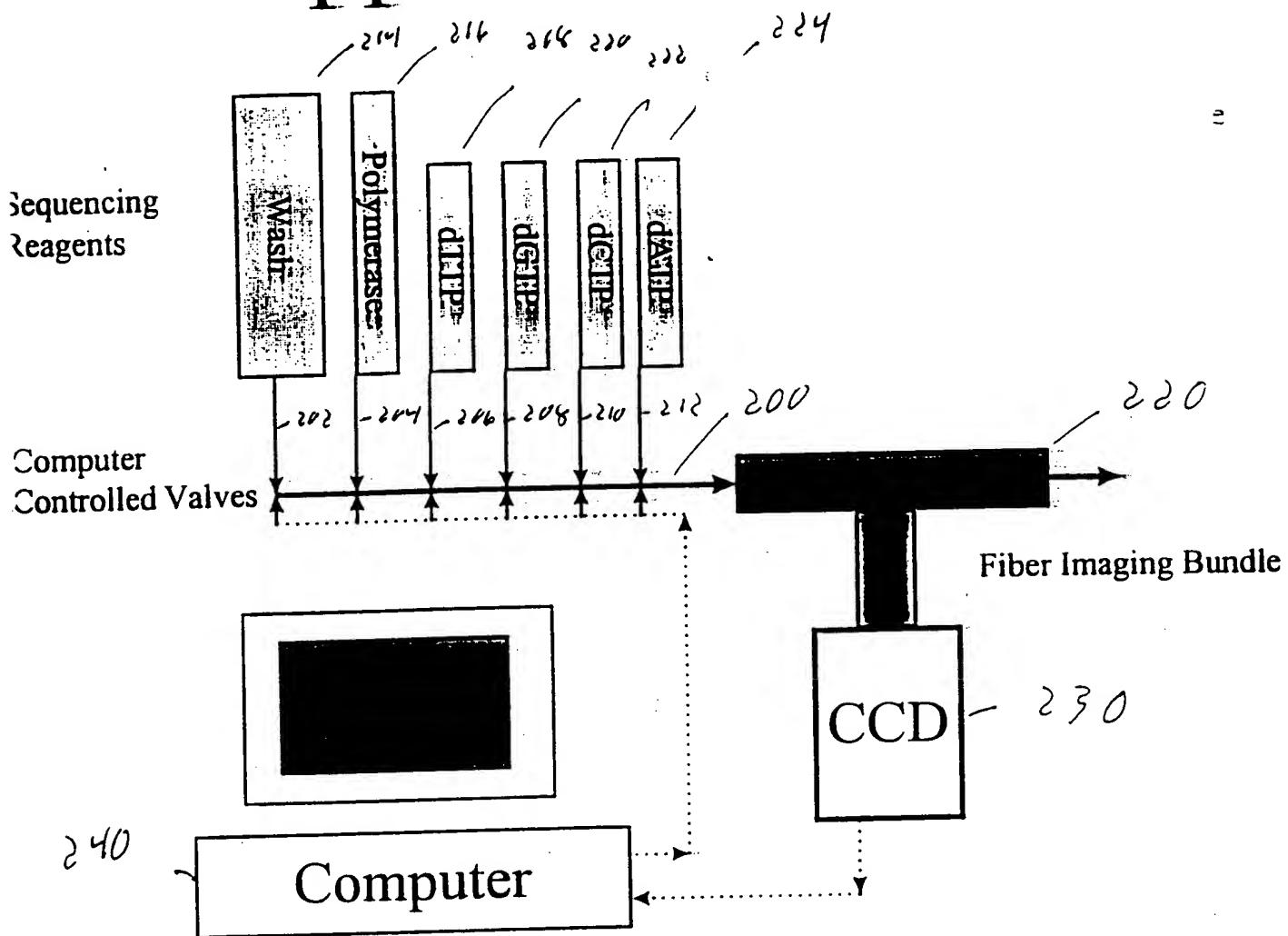


FIG. 1

Apparatus Overview



BEST AVAILABLE COPY

FIG. 2

Perfusion Chamber

Chamber

Chamber Platform

Cover Slip

Top View
Secondary Inlet Port
or Air Trap Release
(RC-21B) Both Sides
Sealant Groove

Perfusion Inlet Port

Bath

Suction Reservoir

Back Pressure Tubing
connects between these two points

Fiber Optic Reactor Array

Side View

Bath Dimensions

Volume 42mm x 42mm

Note: Perfusion chambers can be
clamped with reusable spacer/chamber or pre-
sealed disposable units.

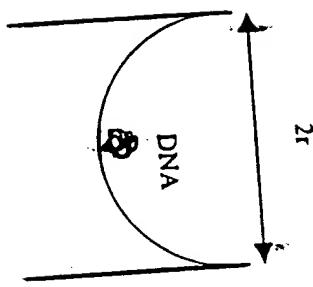


FIG. 4